

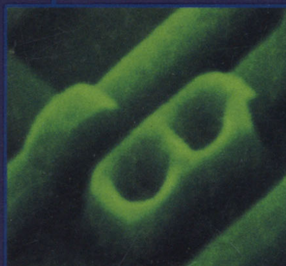
国际信息工程先进技术译丛

纳米CMOS电路和物理设计

Nano-CMOS Circuit and Physical Design

（美） BAN P.WONG ANURAG MITTAL YU CAO GREG STARR 著

辛维平 刘伟峰 戴显英 等译







机械工业出版社
CHINA MACHINE PRESS



VISIT…

LANZAROTE
Caliente.com

关于本书

新技术的快速发展与纳米级缩小的挑战正促使原先相互分离的电路设计、工艺技术、器件物理和物理实现组合在一起形成一门学科。深刻地理解器件、互连和组装方面包含的物理限制对设计电路系统和器件以及做出合理的技术决定都是非常重要的。

《纳米CMOS电路和物理设计》一书将纳米工艺、器件可制造性、先进电路设计和相关物理实现整合到一起，形成了一套先进的半导体技术。这本内容广泛的书探讨了器件和工艺的新发展；提供了设计考虑，重点关注了技术与设计的相互影响，如信号完整性和互连；并且描述了可制造性设计和波动性的影响。重要的主题包括：

- 纳米CMOS工艺缩小问题及其对设计的影响
- 亚波长光刻
- 运行问题的物理与理论以及解决方案
- 可制造性设计和波动性

《纳米COMS电路和物理设计》对集成电路设计者和该领域的专业人员是一本非常有用的书，给他们提供了实际的设计方案和方法。



国际信息工程先进技术译丛

纳米 CMOS
电路和物理设计

（美）BAN P. WONG　ANURAG MITTAL
YU CAO　GREG STARR　著
辛维平　刘伟峰　戴显英　李玉山　史江一　刘　毅　译
贾新章　校



机 械 工 业 出 版 社

本书将纳米工艺、器件可制造性、先进电路设计和相关物理实现等内容整合到一起，形成了一套先进的半导体技术，探讨了器件和工艺的新发展，提供了设计考虑，重点关注了技术与设计的相互影响，并且描述了可制造性设计和波动性的影响。重要的主题包括纳米 CMOS 工艺缩小问题及其对设计的影响；亚波长光刻；运行问题的物理与理论以及解决方案；可制造性设计和波动性。

本书适合集成电路设计者和该领域的专业人员阅读。

Copyright© 2005 by John Wiley & Sons, Inc. , Hoboken, New Jersey.
All Right Reserved. This translation published under license.
本书中文简体字版由 Wiley 授权机械工业出版社独家出版。未经出版者书面允许，本书的任何部分不得以任何方式复制或抄袭。版权所有，翻印必究。

本书版权登记号：图字 01-2008-4175 号

图书在版编目（CIP）数据

纳米 CMOS 电路和物理设计/(美) 王班（Wong, B. P.）等著；辛维平等译. —北京：机械工业出版社，2011. 2
(国际信息工程先进技术译丛)
Nano-CMOS Circuit and Physical Design
ISBN 978-7-111-33083-7

Ⅰ. ①纳…　Ⅱ. ①王…②辛…　Ⅲ. ①纳米材料 - 互补 MOS 集成电路 - 集成电路 - 电路设计　Ⅳ. ①TN432

中国版本图书馆 CIP 数据核字（2011）第 008772 号

机械工业出版社（北京市百万庄大街 22 号　邮政编码 100037）
策划编辑：吕　潇　责任编辑：吕　潇
版式设计：霍永明　责任校对：陈延翔
封面设计：马精明　责任印制：李　妍
北京振兴源印务有限公司印刷
2011 年 4 月第 1 版第 1 次印刷
169mm × 239mm · 22. 5 印张 · 436 千字
0001—3000 册
标准书号：ISBN 978-7-111-33083-7
定价：98.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

电话服务
社服务中心：(010)88361066
销 售 一 部：(010)68326294
销 售 二 部：(010)88379649
读者购书热线：(010)88379203

网络服务
门户网：http：//www. cmpbook. com
教材网：http：//www. cmpedu. com
封面无防伪标均为盗版

译 者 的 话

随着技术的进步，以 CMOS 技术为主流的集成电路已按照摩尔定律预言的发展规律进入到了纳米时代。CMOS 电路在低成本、低功耗、高集成度等方面的明显优势，以及在速度方面的不断提高，使 CMOS 技术已经成为集成电路芯片，特别是 SoC 芯片设计者的最佳选择。纳米 CMOS 技术不仅对工艺技术提出了一系列需要解决的难题，也给电路设计、版图集成以及工艺工程师带来许多新的挑战，要求他们打破相互之间的隔绝状态，进一步把握电路、版图以及工艺之间的相互影响。例如，工艺技术快速发展带来的泄漏增大、工艺波动性的影响增强等问题要求电路与版图设计工程师采取新的对策；为了成功地实现电路功能的不断增加，需要工艺工程师更多地了解电路设计以及版图设计中的一些细节。

针对纳米时代 CMOS 技术的特点，本书的价值在于第一次在同一本书中将工艺与电路、版图设计综合到一起，结合大量的数据与图片资料详细论述了工艺制造与电路、版图设计之间的相互影响，同时介绍了目前最新的技术对策和发展水平。因此本书同时适合集成电路设计和工艺人员阅读。通过本书的学习，可以使设计者对纳米 CMOS 器件和工艺的现状和未来有较全面的了解，对设计者的工作起到指导作用。同时，本书也是纳米 CMOS 工艺工程师了解电路与版图设计的一个窗口。对于大学研究生来说，本书也是一本较好的参考书。

本书第 1 章、第 2 章、第 5 章、第 9 章、第 10 章和第 11 章由辛维平翻译，第 3 章由戴显英翻译，第 4 章和第 6 章由史江一翻译，第 7 章以及第 9 章和第 11 章中关于 RAM 的部分内容由刘伟峰翻译，第 8 章由李玉山翻译，刘毅协助翻译了第 11 章中关于数字电路的部分内容，徐建强对翻译初稿的改善提出了很多修改建议。全书由贾新章审校、统稿。由于本书内容新颖，涉及不少新技术，加之译校者专业和英语水平有限，译文中不当之处甚至错误在所难免，欢迎读者提出宝贵意见，展开讨论。

译、校者　于西安电子科技大学

原　书　序

在过去的数十年中，CMOS 前沿技术不断地发生着翻天覆地的突破性变化，创造了技术上的奇迹。由制造商耗费数十亿美元创建的工厂制造出的线宽小于100nm、集成度达到10 亿个晶体管的复杂集成电路，极大地改变了我们生活的世界。只有许多精通其领域技艺的专家经过不懈努力并发挥他们的聪明才智才可能实现微电子领域的革命。

IC 设计者、器件集成者和工艺工程师已经认识到了广泛理解 IC 技术各个方面带来的益处，正通过持之以恒的学习涉猎相关领域的知识。对 IC 设计者来说，为了充分获得可实现的产品价值，深刻地理解器件、互连和制造方面所包含的物理限制是非常重要的。对于技术开发者，了解技术对先进设计的影响是做出合理技术决定的必要基础。

虽然获取相邻领域知识的需求总是存在的，但是近年来，基于多种原因，使得这种需求更加迫切。新技术引入的步伐与电路速度提升的速率明显快于前 20年。这种加快的步伐可能会延续很长时间，也可能不会持续太久，不过，现在有比以前更大量的新知识需要工程师们去学习与使用。还有一个原因是，随着工艺尺寸的缩小变得越来越困难，诸如泄漏与性能之间、线宽与波动性之间的折衷必须基于对设计、器件和组装的细心考虑，采取比以前更加审慎的方式进行确定。最后一点是大量且越来越多数的工程师工作在专门从事设计或者组装的公司(也就是那些没有芯片制造设备或硅代工线的公司)。这些工程师要面对各种问题，比那些为集成 IC 公司工作的工程师面临更大的挑战。

现在有许多书专注于硅工艺技术或 IC 设计，但是很少能对当前这两个方面的状态进行全面的综览。本书是对纳米工艺、器件可制造性、先进电路设计和相关的物理实现等方面的综合分析，使其更有价值。本书第一部分用 3 章的篇幅对器件和工艺的现状及其未来趋势进行了介绍，第二部分用 6 章的篇幅讲述了设计问题，着重于技术和设计的相互影响，例如信号完整性和互连，以及实际的解决方案。第三部分论述了设计对成品率或可制造性设计的影响。

本书由业界专家撰写，适用于那些需要一本方便而且反映最新发展水平参考书的 IC 设计人员和工艺人员。然而，IC 技术中依然有许多需要开拓的新领域和需要发现的新世界。本书是又一本值得我们随身携带的好书。

CHENMING HU
中国台湾半导体制造公司及加州大学伯克利分校
2004 年 1 月

原 书 前 言

1965 年，戈登·摩尔提出了著名的摩尔定律，并成为半导体工业发展的催化剂。今天，如我们所见，半导体工业前进的巨大动力将我们带到了 sub-100nm 时代。这些进步给工艺控制带来许多困难，并随之对电路与物理设计方面也提出了挑战。因此，设计方法的自由度受到了极大的限制，并且对于将芯片集成在一起的途径方面要求进行革命性的变革，使得集成的芯片不仅具有规定的功能，而且要满足设计目标与高成品率要求。

然而，在缺乏制造设施情况下开发的半导体模型的爆炸性增长，导致工艺/器件工程师与电路设计工程师相互隔绝，而且由于受到工艺和器件物理基础的限制，使得电路设计工程师对于自己的设计给可制造性、成品率和性能带来的影响缺乏了解。当我们进入纳米时代时，掌握如何处理这些问题，对产品乃至公司而言都是至关重要的。这几类工程师必须协同工作以弥补彼此知识的欠缺，并且当我们沿着这条工艺尺寸缩小之路前进时，这种知识的欠缺还在不断扩大，只有协同工作才能实现目标。

当我们在工作过程中面对这些问题的时候，我们发现还没有单独的一本书讨论所有这些问题。这些信息以零碎的形式存在，并且大部分存在于专家们的大脑中，其中一部分内容我们已经在工作过程中得到了解。本书尝试着将这些问题有机地组合成一个整体，讨论它们之间的相互作用以及它们对可制造性、成品率以及性能的影响，给设计者提供实用的指导并帮助设计者克服先进半导体工艺给设计带来的一些困难，同时在物理和电路设计与制造工序、可制造性和成品率之间架设一座迫切需要的桥梁。本书中我们提出的一些概念极其重要，特别是随着工艺不断向纳米 CMOS 特征尺寸技术发展，这些概念的重要性更加显现出来。

本书分为三个部分。第一部分详细描述了深亚微米工艺，以帮助设计者理解与之相关的问题，并且使他们深入了解尺寸缩小带来的限制。第二部分总结分析了工艺缩小对电路设计和物理实现的影响。最后一部分主要着重于可制造性和成品率方面的问题，进而提供指导，以确保设计的器件是可制造的，并且满足成品率和性能目标。

第 1 章总结分析了在深亚微米工艺中设计者所面对的问题，并介绍了本书其余部分的架构。第一部分还包括第 2、3 两章。第 2 章总结分析了 90nm 以及更小尺寸工艺中（管芯工序）（FEOL）和（组装工序）（BEOL）工序系统的当前状态并展望了未来可能的解决方案。FEOL 部分阐述了栅介质和应变工程的发展，包

括相关的设备问题，并且深入讨论了 CMOS 尺寸缩小中遇到的问题，如栅隧穿和 NBTI 等。BEOL 部分讨论了局部和全局互连缩小、铜互连线开发、低 k 介质的挑战、以及集成方案（如双镶嵌工艺）。第 3 章是从物理与理论基础方面介绍光刻的内容，包括了与先进工艺和相应解决方案有关的问题。

第二部分包括第 4 ~9 章，共 6 章。第 4 章简要总结分析了混合信号电路面临的设计问题，并对于如何克服先进工艺下电路设计所遇到的困难提供指导。第 5 章分析了设计者在一个芯片上构建一个复杂系统时面临的 ESD 问题。对于多电源保护一类问题也进行了详细的阐述，从而使设计者具备对特殊的 ESD 要求进行估计的能力。作为发展中的另一种 ESD 保护策略，对最新的 SCR 结构也进行了介绍。第 6 章介绍了 I/O 缓冲器设计的最新发展趋势，并且总结分析了各种 I/O 规范以及当前在设计实现方面的发展趋势。对电源总线问题和同步开关噪声问题进行了详细的讨论，说明了预先开发 I/O 电源总线方案的重要性。另外，还详细的讨论了对晶片上的去耦问题，因为片上去耦问题正成为满足高速接口规范的一个关键问题。第 7 章介绍了 DRAM 设计基础，然后进一步说明了在纳米 CMOS 工艺下成功地实现存储电容、存取晶体管和读出放大器等比例缩小所涉及的相关技术。第 8 章着重于信号完整性分析和片上互连的设计解决方案。首先，介绍了寄生参数的高效提取技术，并特别强调电感的问题。接着讨论了信号时序、串扰噪声和波形完整性分析的分析方法。本章最后探讨了改善高速信号的信号完整性的物理和电路设计解决方案。第 9 章对现有系统设计的各个级别在设计阶段和运行阶段采用的低功耗设计技术进行了全面的讨论，着重阐述了电路级的逻辑和存储器设计方法。第 9 章的最后对未来 90nm 工艺节点以后的超低功耗设计技术进行了展望。

第三部分包括第 10、11 章。第 10 章对实现可制造性设计给出了指导，并给出了多个实例，包括 OPC 后仿真，说明了将电路转换为物理版图时可能出现的问题以及改善的方法。第 11 章介绍了无论工艺如何变化，实现鲁棒且高性能设计的一些原则。本章首先分析了来自工艺和其他方面的各种波动，并讨论了这些波动对电路功能和性能的影响。选用三个主要设计领域（时钟、SRAM 和选定的数字电路）作为案例来说明这些原则。本章还包括了易于进行可制造性设计的指导原则。本章最后简要分析了在纳米 CMOS 设计中对器件统计模型的需求，并简要描述了 BSIM4 模型所具有的新特点。

致谢

我们要对许多为本书的完成做出贡献的人们表示感谢。首先，我们要感谢那些撰写了本书部分章节的学科专家们。我们感谢应用材料公司的工艺专家们——Reza Arghavani、Faran Nouri 与 Gary Miner 先生——感谢他们对前道工序的设备

要求部分所做的贡献。我们要感谢应用材料公司的 Khaled Ahmad 先生，他提供了第 2 章前道工艺部分采用的氧化层特性方面的数据。我们要感谢加利福尼亚大学伯克利分校的工艺专家 Qiang Lu 先生，他目前就职于 AMD 公司，感谢他对 FEOL 章节的贡献。我们也要感谢 IBM 微电子的光刻专家 Franz Zach 先生，感谢他关于第 3 章中纳米 CMOS 阶段的光刻技术介绍。对第 5 章，我们要感谢被公认为这方面主题的权威——中国台湾大学的 Ming-Dou Ker 教授。我们要感谢英飞凌公司的存储器权威 Martin Brox 先生对第 7 章的贡献。我们要感谢 Rambus 的 Xuejue Huang 先生对第 8 章的卓越贡献，还要感谢 UC-Berkeley 的 Huifang Qin 先生，他撰写了第 9 章的大部分内容，并且将作者的工作整合到本章中。

我们还要感谢 Altera 公司对本书的支持，特别是 Wanli Chang、William Hwang、KangWei Lai、Richard Chang、Leon Zheng、Mian Smith 和 Howard Kahn 先生，感谢他们的仿真工作。我们感谢 Cynthia P. Tran 先生提供了作为本书中插图的物理版图以及光刻仿真的输入。我们感谢 John Madok 和 Michael Smayling 先生，他们协助推荐应用材料的专家撰写本书中的部分章节并担任顾问。

我们十分感谢 Trecenti/Hitachi 公司的 Shuji Ikeda 先生，NEC 公司的 Ryuichi Hashishita、Yashushi Yamagata 和 Toshiaki Hoshi 先生；AMD 公司的 Richard Klein 和 Qiang Lu 先生，感谢他们提供了本书中使用的技术数据和多幅 SE 和 TE 显微照片。我们感谢 ASML Masktools 公司的 Fung Chen、Armin Liebchen 和 Sabita Roy 先生，感谢他们在光刻仿真以及提供仿真工具用来产生作为插图的仿真光刻胶剖面的俯视图。

我们感谢英国哥伦比亚大学的 Mark Greenstreet 教授，他审阅了本书的编写提纲并给出了许多有价值的建议。最后，我们表示对 Chenming Hu 教授的感谢，感谢他提出的卓有远见的建议以及为本书撰写的序。

作　者

目　　录

译者的话
原书序
原书前言
第1章　纳米 CMOS 的缩小问题及内涵 …… 1
1.1　纳米 CMOS 时代的设计方法 …… 1
1.2　使得性能改善得到延续所必需的创新 …… 3
1.3　sub-100nm 缩小的挑战和亚波长光刻综述 …… 5
1.3.1　后道工艺的挑战（金属化） …… 5
1.3.2　前道工艺的挑战（晶体管） …… 10
1.4　工艺控制和可靠性 …… 13
1.5　光刻问题和掩膜数据爆炸 …… 14
1.6　新型的电路和物理设计工程师 …… 15
1.7　建模的挑战 …… 16
1.8　变革设计方法的需要 …… 17
1.9　总结 …… 19
参考文献 …… 19
第2章　CMOS 器件与工艺技术 …… 22
2.1　前道工序的设备要求 …… 22
2.1.1　技术背景 …… 22
2.1.2　栅介质的缩小 …… 24
2.1.3　应变工程 …… 28
2.1.4　快速热处理技术 …… 30
2.2　在 CMOS 尺寸缩小中与前道工序相关的器件问题 …… 36
2.2.1　CMOS 缩小的挑战 …… 36
2.2.2　量子效应模型 …… 38
2.2.3　多晶硅栅耗尽效应 …… 40
2.2.4　金属栅电极 …… 42
2.2.5　栅直接隧穿泄漏电流 …… 43
2.2.6　寄生电容 …… 45
2.2.7　需要关注的可靠性问题 …… 48
2.3　后道工序互连线技术 …… 50

2.3.1　互连缩放 …… 51
2.3.2　铜互连技术 …… 52
2.3.3　低 k 介质的挑战 …… 54
2.3.4　未来的全局互连技术 …… 55
参考文献 …… 56
第 3 章　亚波长光刻的理论与实践 …… 63
3.1　引言与成像理论概述 …… 63
3.2　对 100nm 节点的挑战 …… 65
3.2.1　100nm 节点的 k 因子 …… 65
3.2.2　重要的工艺波动 …… 67
3.2.3　低 k 成像对工艺敏感性的影响 …… 70
3.2.4　低 k 成像和聚焦深度的影响 …… 71
3.2.5　低 k 成像和曝光容限 …… 71
3.2.6　低 k 成像及其对掩膜误差增强因子的影响 …… 72
3.2.7　低 k 成像及其对像差的敏感性 …… 73
3.2.8　低 k 成像以及 CD 变化与条宽的关系 …… 74
3.2.9　低 k 成像和拐角处的圆角半径 …… 76
3.3　分辨率增强技术：物理 …… 78
3.3.1　专门的照明模式 …… 79
3.3.2　光学邻近修正（OPC） …… 80
3.3.3　亚分辨率辅助图形 …… 86
3.3.4　交替式相移掩膜 …… 88
3.4　物理设计风格对 RET 和 OPC 复杂性的影响 …… 91
3.4.1　特定照明条件 …… 92
3.4.2　二维版图 …… 94
3.4.3　交替式相移掩膜 …… 98
3.4.4　掩膜版成本 …… 101
3.5　发展前景：未来的光刻技术 …… 103
3.5.1　发展之路：157nm 光刻 …… 103
3.5.2　进一步演化：浸没式光刻 …… 104
3.5.3　巨大突破：EUV 光刻 …… 106
3.5.4　粒子束光刻 …… 107
3.5.5　直写式电子束设备 …… 108
参考文献 …… 111
第 4 章　混合信号电路设计 …… 115
4.1　引言 …… 115

4.2 设计考虑 …… 115
4.3 器件建模 …… 116
4.4 无源器件 …… 122
4.5 设计方法学 …… 125
4.5.1 工艺测试基准电路 …… 126
4.5.2 薄氧器件设计 …… 126
4.5.3 厚氧器件设计 …… 127
4.6 低压技术 …… 129
4.6.1 电流镜 …… 129
4.6.2 输入级 …… 131
4.6.3 输出级 …… 132
4.6.4 带隙基准 …… 132
4.7 设计过程 …… 133
4.8 静电放电保护 …… 135
4.8.1 针对多电源情况的考虑 …… 136
4.9 噪声隔离 …… 137
4.9.1 保护环结构 …… 137
4.9.2 隔离的 NMOS 器件 …… 139
4.9.3 外延材料与体硅 …… 139
4.10 去耦 …… 140
4.11 主电源线 …… 144
4.12 集成问题 …… 144
4.12.1 芯片四角区域的影响 …… 144
4.12.2 邻近电路的影响 …… 145
4.13 总结 …… 145
参考文献 …… 146
第 5 章 静电放电保护设计 …… 149
5.1 引言 …… 149
5.2 ESD 标准与模型 …… 149
5.3 ESD 保护设计 …… 150
5.3.1 ESD 保护方案 …… 150
5.3.2 ESD 保护器件开启的一致性 …… 152
5.3.3 ESD 注入与金属硅化物阻挡 …… 153
5.3.4 ESD 保护指导意见 …… 154
5.4 针对高速 I/O 的低电容 ESD 保护设计 …… 154
5.4.1 高速 I/O 或模拟引脚的 ESD 保护 …… 154

5.4.2 小电容ESD保护设计 …… 156
5.4.3 输入电容的计算 …… 159
5.4.4 ESD鲁棒性 …… 160
5.4.5 开启验证 …… 161
5.5 混合电压I/O的ESD保护设计 …… 165
5.5.1 混合电压I/O接口 …… 165
5.5.2 混合电压I/O接口的ESD问题 …… 165
5.5.3 混合电压I/O接口的ESD保护器件 …… 167
5.5.4 混合电压I/O接口的ESD保护电路设计 …… 170
5.5.5 ESD鲁棒性 …… 172
5.5.6 开启验证 …… 173
5.6 用于ESD保护的SCR器件 …… 174
5.6.1 SCR器件的开启机制 …… 175
5.6.2 基于SCR的CMOS片上ESD保护器件 …… 176
5.6.3 SCR闩锁工程 …… 183
5.7 总结 …… 185
参考文献 …… 186
第6章 输入/输出设计 …… 193
6.1 引言 …… 193
6.2 I/O标准 …… 194
6.3 信号传输 …… 195
6.3.1 单端缓冲器 …… 195
6.3.2 差分缓冲器 …… 196
6.4 ESD保护 …… 199
6.5 I/O开关噪声 …… 200
6.6 匹配 …… 203
6.7 阻抗匹配 …… 206
6.8 前置放大 …… 206
6.9 均衡化 …… 208
6.10 总结 …… 209
参考文献 …… 210
第7章 DRAM …… 212
7.1 引言 …… 212
7.2 DRAM基础 …… 212
7.3 电容的缩放 …… 215
7.4 阵列晶体管的缩放 …… 217

7.5 读出放大器的缩放 …… 220
7.6 总结 …… 223
参考文献…… 223
第8章 片上互连的信号完整性问题…… 225
8.1 引言 …… 225
8.1.1 互连的品质因数…… 227
8.2 互连参数提取 …… 228
8.2.1 互连的等效电路表示 …… 229
8.2.2 *RC* 的提取 …… 232
8.2.3 电感提取 …… 235
8.3 信号完整性分析 …… 239
8.3.1 互连驱动器模型…… 239
8.3.2 *RC* 互连分析 …… 241
8.3.3 *RLC* 互连分析 …… 244
8.3.4 考虑噪声耦合效应的时序分析 …… 247
8.4 信号完整性设计技术 …… 249
8.4.1 物理设计技术 …… 250
8.4.2 电路技术 …… 254
8.5 总结 …… 258
参考文献…… 259
第9章 超低功耗电路设计…… 263
9.1 引言 …… 263
9.2 设计阶段的低功耗技术 …… 264
9.2.1 系统级和结构级设计阶段的低功耗技术 …… 264
9.2.2 电路级设计阶段的低功耗技术 …… 265
9.2.3 设计阶段的存储器技术 …… 269
9.3 运行阶段的低功耗技术 …… 274
9.3.1 运行阶段的系统级和结构级低功耗技术 …… 274
9.3.2 针对运行阶段的电路级低功耗技术…… 277
9.3.3 针对运行阶段的存储器低功耗技术…… 279
9.4 低功耗设计的技术革新 …… 283
9.4.1 新颖的器件技术…… 283
9.4.2 组装技术革新 …… 284
9.5 未来超低功耗设计的展望 …… 285
9.5.1 亚阈区电路工作…… 285
9.5.2 容错设计 …… 286

9.5.3　异步设计与同步设计 …… 286
9.5.4　栅感应泄漏抑制方法 …… 286
参考文献 …… 287
第10章　可制造性设计 …… 294
10.1　引言 …… 294
10.2　最优和亚最优版图对比 …… 295
10.3　全局布线 DFM …… 300
10.4　模拟电路的 DFM …… 301
10.5　一些基本规则 …… 303
10.6　总结 …… 304
参考文献 …… 304
第11章　针对波动性的设计 …… 305
11.1　波动性对未来设计的影响 …… 305
11.1.1　电路设计中的参数波动 …… 305
11.1.2　对电路性能的影响 …… 307
11.2　减轻波动影响的策略 …… 309
11.2.1　使偏斜最小化的时钟分布策略 …… 309
11.2.2　针对波动性的 SRAM 技术 …… 312
11.2.3　应对波动性的模拟电路策略 …… 321
11.2.4　应对波动的数字电路策略 …… 329
11.3　纳米 CMOS 工艺角建模方法 …… 335
11.3.1　统计模型的需求 …… 335
11.3.2　统计模型的使用 …… 336
11.4　BSIM4 模型的新特点 …… 340
11.4.1　halo/packet 注入 …… 340
11.4.2　栅感应漏极泄漏和栅直接隧穿 …… 341
11.4.3　建模的挑战 …… 342
11.4.4　与建模相关的问题 …… 343
11.4.5　模型总结 …… 343
11.5　总结 …… 343
参考文献 …… 343

第 1 章　纳米 CMOS 的缩小问题及内涵

1.1　纳米 CMOS 时代的设计方法

随着工艺特征尺寸缩小到 100nm 节点以下，对于硅集成电路，为了满足功能和高成品率的要求，需要针对工艺波动的增加、互连工艺的困难和其他一些新出现的更严重的物理效应，改变传统的设计方法。纳米 CMOS 中栅氧的减薄（见图1-1）致使栅直接隧穿电流急剧增加。亚阈区泄漏和栅直接隧穿电流（见图 1-2）不再是二级效应了[1,15]。设计中将要考虑栅感应漏极泄漏（Gate-Induced Drain Leakage，GIDL）的影响。例如对于 DRAM（第 7 章）和低功耗 SRAM（第 9 章），它们的栅压相对于源极来说是负电位[15]，如果不考虑这些效应，得到的 SRAM、DRAM 或者任何其他使用这个技术来减小亚阈区泄漏的电路会丧失功能。某些情况下甚至多路器和触发器也会受到影响。

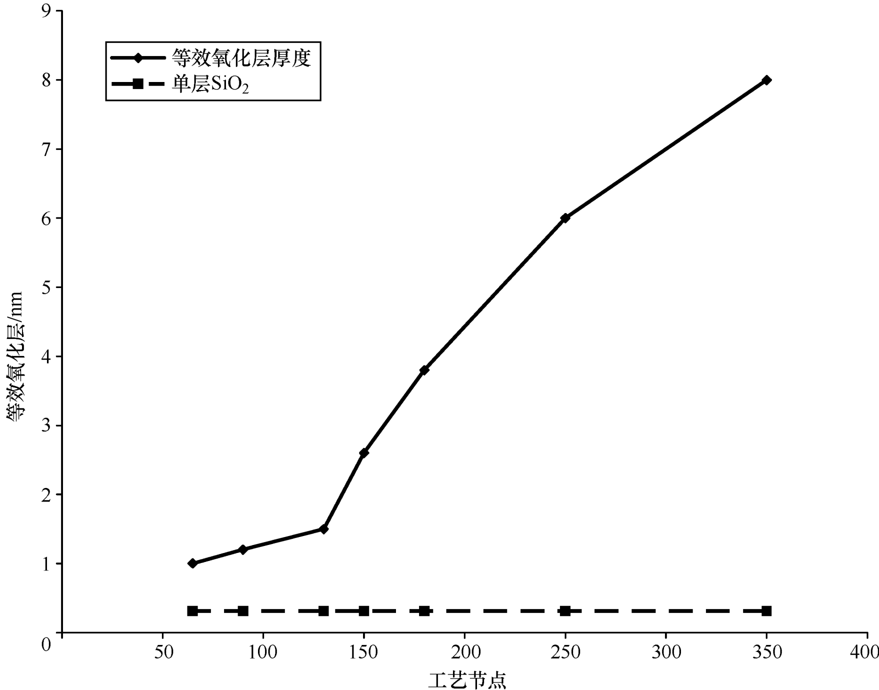


图 1-1　栅氧厚度随工艺节点的变化趋势

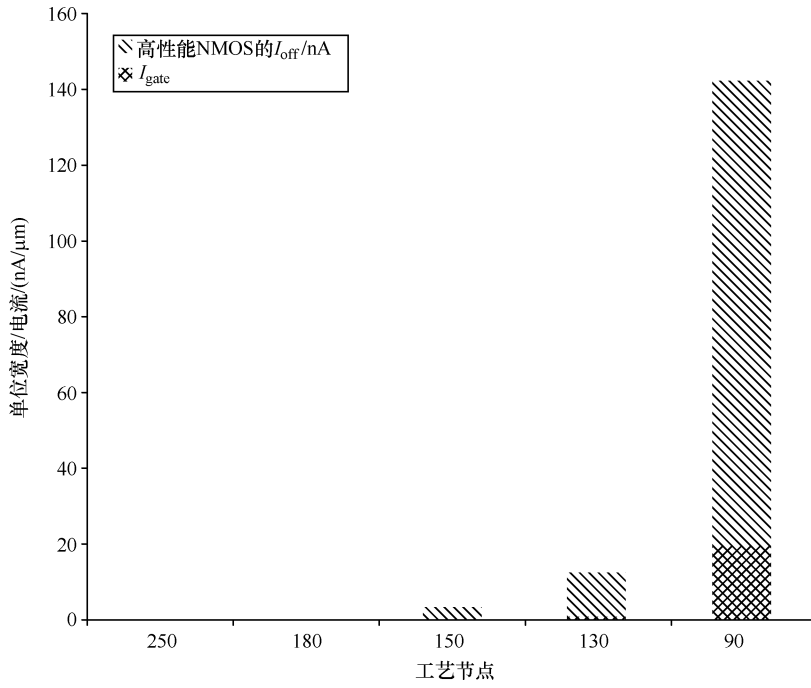


图 1-2 I_{gate}和亚阈区泄漏随工艺节点的变化趋势

在功能级需要面对的问题不仅仅是亚阈区泄漏和栅电流，还需要应对高性能电路，如微处理器、数字信号处理器和图形处理单元等芯片中的电源管理问题。电源管理也是移动应用中需要面对的一个挑战。

另外，即使采用增强分辨率扩展技术（Resolution Extension Technology，RET），光刻也会趋于极限。这些新的光刻技术还会使掩膜版费用增加，并且使生产周期更长。经过几次修改才能得到一个值得进行生产的设计，这将明显增加研制费用。过去，处理器设计者一般在验证置信度达到 98% 时就可以提交设计数据，在随后的工艺过程中还可以继续排错，这样一般能够使产品的研制速度快几个数量级，从而使产品投放到市场的时间更快。现在，由于掩膜费用的增加和制造周期变得更长，在获得最低的研制成本和以最短的时间将产品投放市场之间进行折衷，将会相当困难[28]。

因为设计规则不能以相同的速率缩小，原先成功的设计在下一代工艺中必须完全重新设计，除非设计者可以预见设计规则的变化，并且在针对原先尺寸较大的工艺技术进行设计时就有意识地牺牲密度以便在下一代工艺技术中可以缩小设计而无需重新设计物理版图。但是仍然必须对关键电路进行再次仿真。如果设计者使用易于缩小的电路技术，也能实现版图的最小化。为此需要预先的计划以及对设计规则进行折衷从而获得一个可以缩小的设计，所以，可以通过工艺缩小以

及很少的重新设计（如果需要的话），采用中等的投入提升性能，获得一个速度更快、面积更小的芯片。预测设计规则变化趋势的关键是要对工艺难点和工具设备的制约具有充分的理解，这将会在以后几章中详细讨论。

1.2　使得性能改善得到延续所必需的创新

晶体管的品质因数现在已经偏离了与栅长倒数之间的关系。如图 1-3 所示，4 扇出延迟随着现代工艺技术进步而变小的趋势在减小。另外，芯片总体互连线并不能缩短，而 0.1μm 以下的互连线电阻正在以指数关系增长。这主要是由于表面散射和窄槽中晶粒尺寸的限制，导致载流子散射和迁移率退化[2]。目前栅介质的厚度已经接近原子尺寸，在 90nm 节点中栅氧化层约为 1.2nm，为五层原子[22]。如图 1-1 所示，栅氧在趋于其一层原子厚度的极限过程中，减薄的趋势正在减缓[26]。源漏扩展电阻（RSD）在晶体管"导通"电阻中占据的比例正不断增大。在 130nm 节点中，源漏扩展掺杂已经增加得非常显著，减小扩展电阻的能力必须与其他短沟效应，如热载流子注入（Hot-Carrier Injection，HCI）以及由能带到能带隧穿引起的泄漏电流，进行折衷。源漏扩散层已变得很薄以致于注入处于饱和水平，电阻不再能减小，除非改用其他掺杂剂[21]。

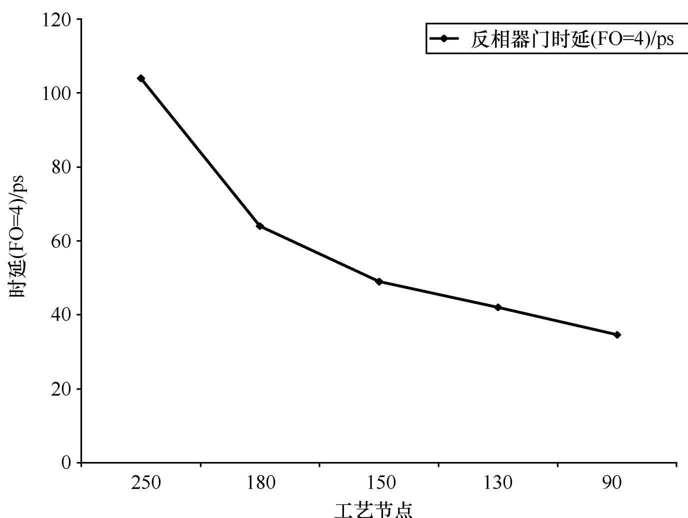


图 1-3　栅延迟与工艺节点的关系

现在的多晶硅连线变得相当窄，130nm 工艺约为 70 ~90nm，而 90nm 工艺为 50nm（见图 1-4）。这要求在多晶硅方阻与源漏泄漏之间进行折衷。为了降低较窄多晶硅连线的电阻，要求多晶硅形成更多的硅化物。因为多晶和源漏扩散都会发生硅化过程，增加多晶硅的硅化物会导致源漏扩散发生更多的硅化消耗。由于

源漏采用超浅结，源漏扩散硅化消耗会导致穿通。在 65nm 工艺中，正在研究在生产过程中采用抬高的源漏技术以缓解这种效应，并且可能用于 90nm 工艺。一些厂商可能将这种在线技术应用到目前建立的 90nm 工艺过程中。

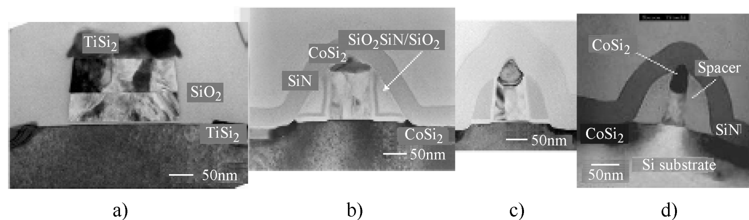


图 1-4 晶体管 TEM [a)、b)、d) 源于 NEC 和 Trecenti/Hitachi；c) 源于 Advanced Micro Devices，Inc.，引用已得到许可]
a) 250nm b) 130nm c) 90nm d) 65nm

从 180nm 工艺开始，特征尺寸（多晶）已经达到亚波长阶段，即特征尺寸小于光刻中使用的紫外线波长（UV）。随着下一代工艺的发展，这种差距正在扩大（见图 1-5）。在 65nm，即使有先进的 RET，193nm 光刻也已经走到了尽头。为了增加 193nm 扫描器的分辨率。正通过研究增加了光刻系统的数值孔径（Numerical Aperture，NA），包括采用浸入式光刻。关于光刻方面所面临的挑战将在第 3 章详细描述。157nm 的挑战和远紫外线（EUV）光刻是新的里程碑，并且会增加设备和掩膜费用以及生产周期。如果 157nm 光刻不能用于 65nm，亚波长差距将进一步扩大。电路和物理设计者将不可能再简单地采用工艺设计规则进行设计，也不可能指望生成一种可缩放的功能设计，只采用一组掩膜版就能满足设计目标可变的要求，例如高性能和低功耗的移动应用。设计者必须明白什么时候采用宽松的规则，

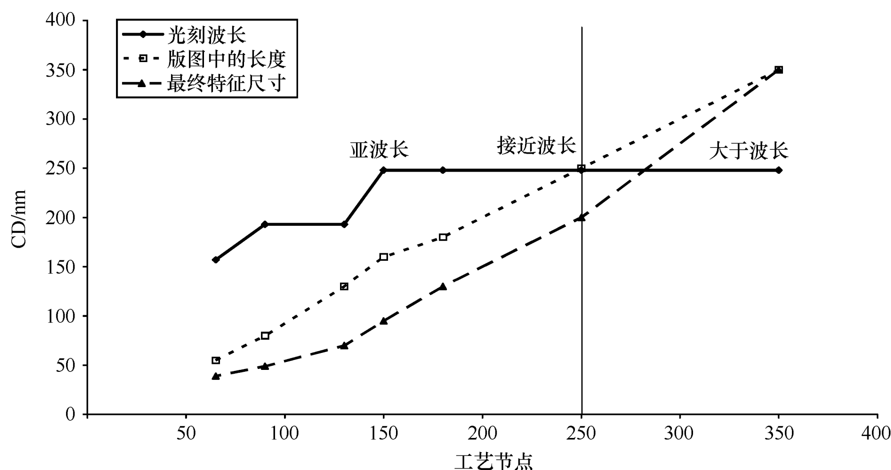


图 1-5 各代工艺节点中的多晶 CD 与光刻 UV 波长

而不是简单地对整个设计放宽规则，否则会给物理缩小带来负面结果。

用来制造新结构的材料和工艺的组合所产生的集成复杂度需要同时从设计和版图多方面考虑解决方案[20]。工艺工程师和技术开发人员不能解决所有这些在 sub-100nm 缩小中出现的问题，这些问题包括集成的复杂度以及制造和工艺控制方面的困难。我们将给电路和物理设计者在应对 sub-100nm 工艺中遇到的困难方面提出可以采用的技术建议，并且帮助他们加深对设计中采用的工艺技术的理解。同样，对工艺工程师来说，理解物理设计基础也非常重要，从而可以改进技术，满足一个可以持续实现物理和性能缩小，并且可缩放的鲁棒设计的需要。

技术设计者需要进行一些革新，在工艺线上采用一些新工艺。开发新材料也是必需的。一个无可争议的事实是，仅由物理缩小获得性能提升的方法已经到达了一个拐点，并且即便可能的话，也不能提供更多的改进。为了继续进行性能提升，在研究工作中已经有一些创新的改进，但更多的是处于开发中。绝缘层上硅（SOI）技术已经证明可以将晶体管的性能提高约 20% ~30%，数值的差别在于数据来源不同。作为技术选择，一些微处理器已经采用 SOI。实验证明，采用宽松的锗化硅衬底的应变硅可以将载流子迁移率提高达 30%。然而，这些衬底材料成本高，并且易于产生位错缺陷，因此还没有被广泛应用。

在载流子迁移率提高方面，还有一种已经证实可以采用的硅上应变方式的革新是使用氮化硅“帽”层。漏源扩散上的压应力在这一层产生应变，由于漏源扩散相互隔开，就在晶体管的沟道中产生应变。但是，它仅适用于 90nm 及其以下工艺，因为沟道必须非常接近漏源应力。长沟器件几乎不会获得好处。即使在 90nm 工艺的晶体管中，如果版图上沟道长度大于最小值，也会使这种获益减小。不幸的是，在 130nm，这种能提高性能的选项受到了限制。这个技术将会是产生应变的首选方式。因为它不要求特殊的衬底，并且迄今为止也未发现有层错。其最大的优点是，不需要增加工艺步骤，只需要调整工艺顺序。

采用铜互连替代铝互连使得因尺寸缩小带来的性能提升不断接近极限所承受的压力获得了短暂的缓解。这是一个革新需要引入新材料的实例。许多其他得到普遍认可的新技术尚处于从研究到试用阶段，包括提高源漏（SD）扩散、双栅 FET、FinFET、高 k 栅介质和金属栅[4]。它们能否最终得到成功应用取决于风险收益比，以及成本、集成与制造复杂度和制造周期。

1.3　sub-100nm 缩小的挑战和亚波长光刻综述

1.3.1　后道工艺的挑战（金属化）

金属电阻　互连线宽度小于 0.1μm 时，电阻率会随着线宽的减小以指数形

式增加。对较窄的互连线来说，高电阻率阻挡层材料占据了导体横截面面积的较大部分。由于表面散射导致的电子迁移率减小成为电阻率增加的因素之一[2]。窄互连中形成的晶粒更小，在填充凹槽时不能再结晶为更大的晶粒，使得电阻率进一步增大。

另外，阻挡层和凹槽（互连线宽度）特征尺寸（Critical Dimension，CD）的变化导致电阻的变化更大。这些因素与化学机械抛光（Chemical-Mechanical Polish，CMP）形成的凹陷和腐蚀，以及光刻和刻蚀变形一起会引起互连线电阻的进一步变化[19]（见图 1-6）。

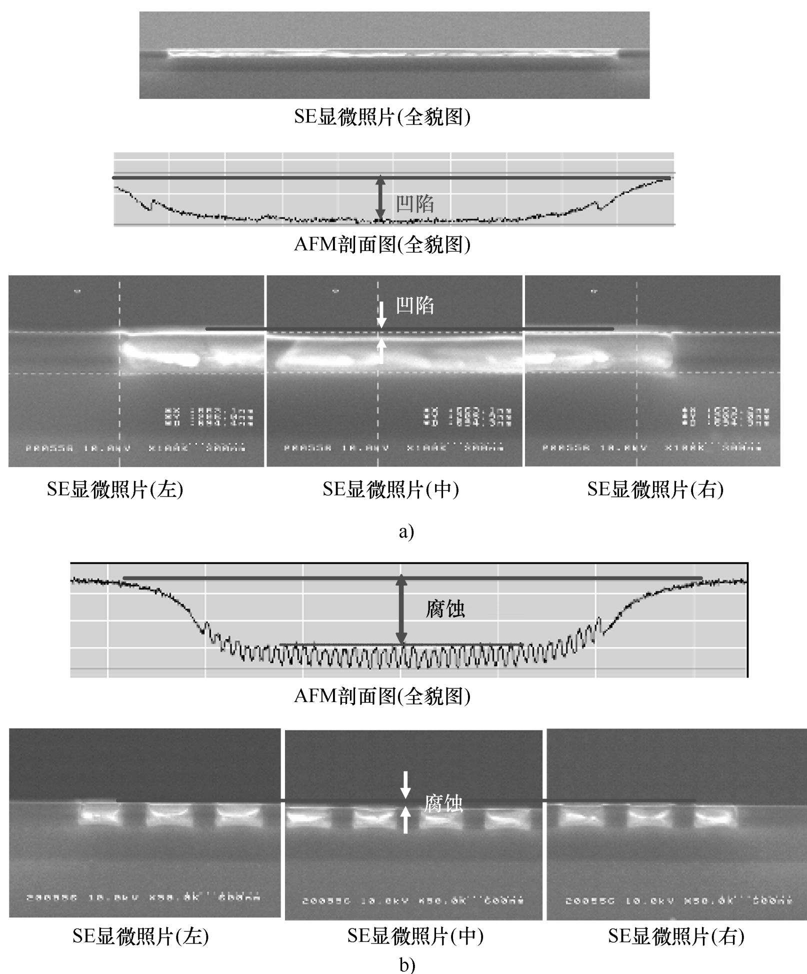


图1-6　a）互连凹陷：较宽互连线的面积　b）互连腐蚀：互连线和间距面积
（图片来源于 Trecenti/Hitachi）

130nm 工艺的互连 *RC* 值不断增加，并且对小于 130nm 的情况，局部和全局互连都变得更差。像上面解释的那样，随着线条尺寸的减小，电阻率不断增加（见第 2 章图 2-25），而电容并未减小。其结果是，即使局部互连线的长度变得越来越短，而局部互连线的时延却在增加（见图 1-7 ~ 图 1-9）。虽然线条尺寸在减小，但是因为新的设计中增加了更多的功能，因此芯片面积并没有减小，全局互连线的长度也就没有减小。例如，180nm 工艺生产的 Pentium 4 Willamette core 有 4200 万个晶体管；而 130nm 工艺的 Northwood core 的晶体管数达到了 5500 万个。这是因为在 Northwood core 中 L2 缓存从 256KB 增加到了 512KB。随着工艺的发展，在一个时钟周期里信号可到达的区域与芯片面积之比正在减小。对一个采用先进工艺的设计来说，时钟频率在增加，而晶片的尺寸并没有减小，使这一问题更加严重。

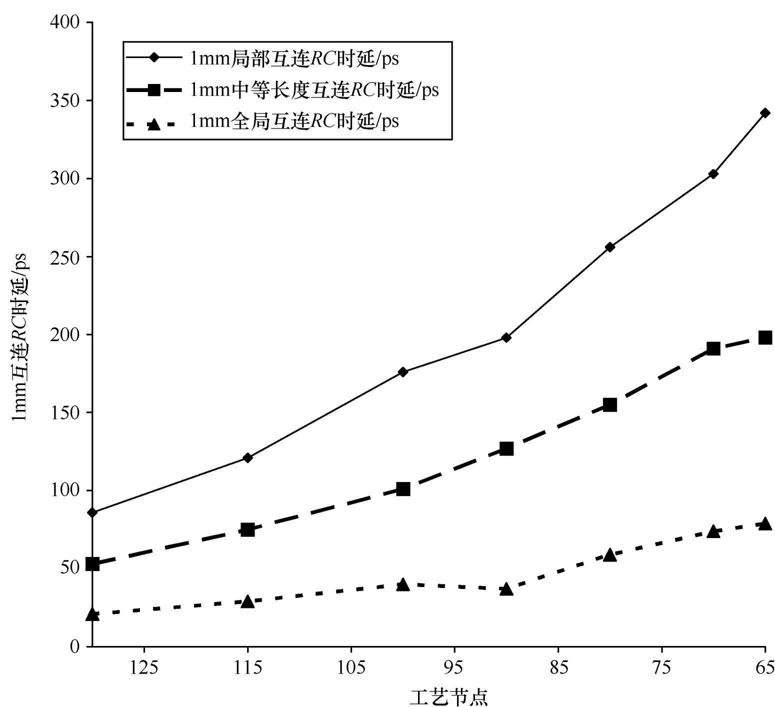


图 1-7　互连时延与工艺节点的关系

互连介电常数　低 k 介质使纳米 CMOS 中互连线缩小成为可能，但是随着互连线宽度和间距不断减小，使得低 k 介质的使用更加困难。这是由于间距较窄互连线之间的经时介质击穿（Time-Dependent Dielectric Breakdown，TDDB），使得低 k 介质存在着潜在的泄漏和可靠性风险。封装工艺带来的问题要求芯片上形成一层“硬外层”，使得管芯具有足够的机械强度，能够承受封装工艺施加给芯片

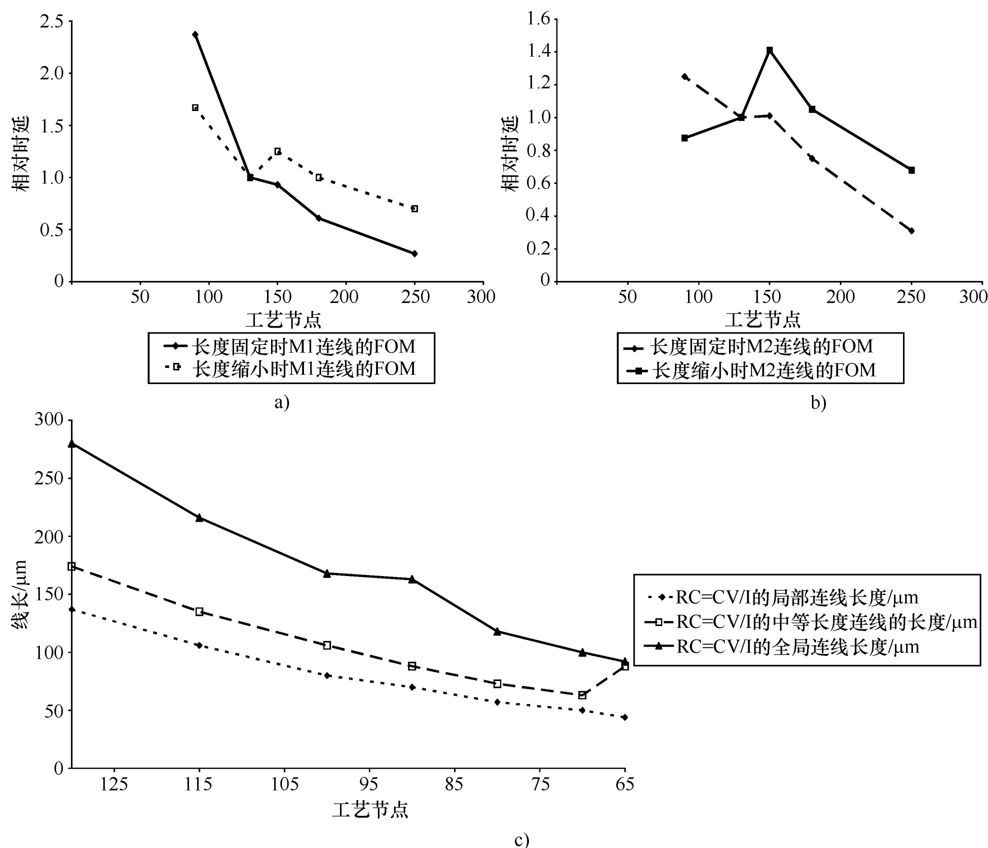


图 1-8 a) M1（局部互连）的品质因数（无 Miller，无重复）
b) 中等长度互连的品质因数（无 Miller，无重复）
c) 等效于 NMOS CV/I 的互连线长度与工艺节点的关系

的应力。这种硬外层的要求意味着位于层叠金属上面的几层需要采用较高介电常数的介质材料，这会稍稍减小低 k 金属技术的有效性。在 8 ~9 层金属的工艺中，低 k 介质将被限制到只能用于 4 ~5 层。具体情况取决于上层金属的使用方法。一般地，上层金属被用来作功率分布。在大多数设计中，它们也被用作时钟分布层，因此就增加了时钟网络的功耗并且也要求更多级去缓冲来自 PLL 的信号，这会导致更高的时钟偏差。

低 k 互连的应用严重缓慢 低 k 技术的引入比较缓慢的原因在于铜互连阻挡材料带来的问题、应对封装过程中应力的机械完整性问题以及制造过程中的一系列问题。这使得几个厂商转而采用氟硅玻璃（FSG）介质。

低 k 介质形似果冻并且多孔，因此很容易受潮、吸附杂质并且气体也容易渗透。因为材料很软，很容易受 CMP 损伤，引起成品率降低和侵蚀，进而影响互

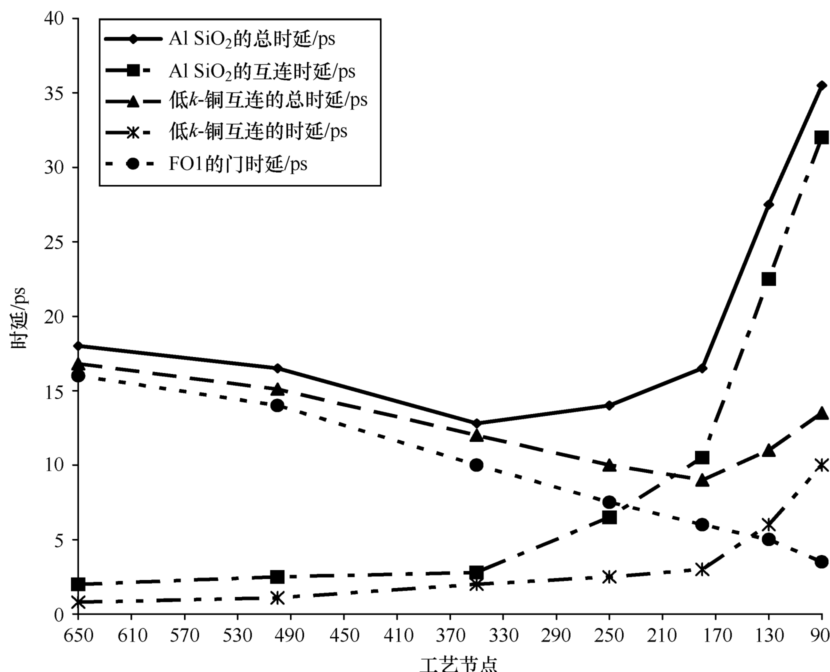


图 1-9　栅时延和互连时延与工艺节点的关系

连线电阻率。低 k 介质的导热性不良，这样会使互连的抗电迁移（EM）特性退化，在一定程度上抵消了一些铜互连 EM 特性好的优点。

互连的品质因数（Figure of Merit，FOM）　对每代工艺节点，按照其设计规则确定的互连线的 FOM 一直在下降（见图 1-7～图 1-9）。过去，晶体管性能滞后。现在，我们已经到达了这样一个转折点，即互连性能已成为芯片性能的制约因素。局部互连性能未得到改善，而全局互连实际上正变得更慢，特别是如果增加了附加的功能使得互连长度没有减少[12-14]。不管工艺技术如何缩小，由于在新的设计中功能不断增加，芯片尺寸却没有减小，总是维持在与以前设计相同的尺寸上。其他情况下，例如微处理器，不管工艺技术如何缩小，芯片尺寸实际上都在增加。尽管工艺技术在缩小，如果芯片变得更大，在不同模块之间就需要全局互连传递信号。

已经预见到的是，当工艺技术缩小时，时钟频率将增加，而一个周期中可达到的区域与总芯片面积的比率会下降[13]。为了使互连主宰的路径可以实现更好的缩小并且不限制频率，就迫使设计者在全局互连上插入更多的中继器，并且有时在全局信号中必须采用流水线方式。然而，这将增加芯片面积、功耗和时钟负载[14]，也增加整个芯片时序的复杂性。更高的时钟负载将导致更高的时钟偏差。由于采用流水线，就增加了信号的等待时间，这对其他的微结构也会产生影响。

这些问题迫使设计者只好放宽节距，以改善全局互连线的性能和信号完整性。增加互连的节距可以减小线间耦合，但是电容将逼近渐进值，即随着互连线之间间距的进一步增大，电容将不再会随之减小（见图 1-10）。使得电容为最小时的间距还与层间介质的厚度有关。工艺技术进一步缩小到小于 130nm 节点时，由于性能问题和信号完整性问题，互连密度几乎不会得到改善，这将要求对一些互连需要采取屏蔽措施，而对另外一些互连需要增大互连之间的间距。这样，对复杂芯片进行布线，会要求采用更多的金属层。

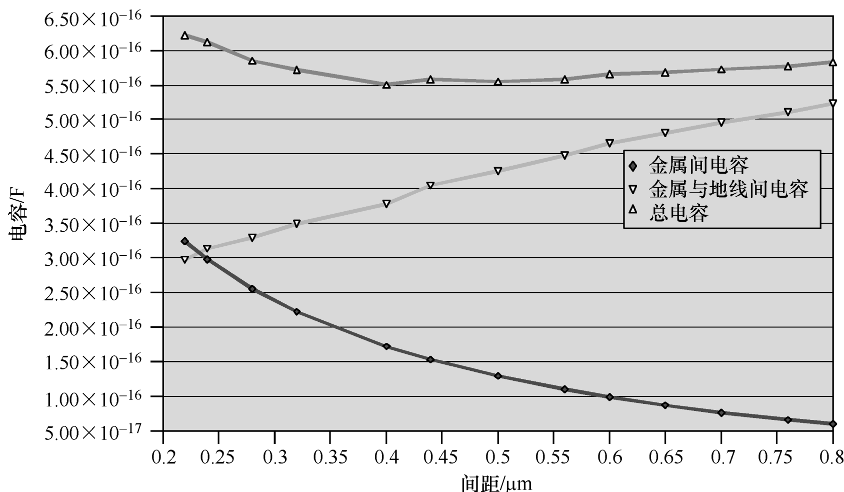


图 1-10　金属-金属电容与间距的关系

接触孔和通孔不能进一步缩小　对大多数 130nm 技术来说，接触孔尺寸已经是 0.16μm，而通孔尺寸是 0.2μm。在未来的工艺中，很难使它们缩减得更多，它们将不能与其他特征尺寸一样以相同的速率缩小。另一个限制因素是接触孔和通孔电阻，随着接触孔和通孔尺寸的缩小，电阻将上升。

130nm 节点中，接触孔和通孔这两层已经要求采用光学邻近修正（Optical Proximity Correction，OPC）和相移光刻。这些层掩膜版的数据以及掩膜版制作费用已经是其他不需要 OPC 和/或相移层次的两倍[5]。

1.3.2　前道工艺的挑战（晶体管）

晶体管性能　晶体管的品质因数现在正偏离与栅长倒数成正比的关系。导致这一结果的主要因素是：

- $V_{gs}-V_{th}$ 不断缩小，而 V_{th}/V_{dd} 越来越大（见图 1-11）。
- 作为晶体管总导通电阻一部分的 RSD 的作用正变得更大，导通电阻部分由多晶栅与接触孔之间的间距以及 RSD 决定。

- 浅结正使得掺杂水平达到饱和。使得 RSD 没有进一步减小的可能；同时，结电容却不断增加。
- 由于电流的聚集，更薄的源和漏扩散使得 RSD 进一步增大。
- 浅沟槽隔离（Shallow Trench Isolation，STI）应力感应的迁移率退化更加显著，虽然 PMOS 晶体管在 STI 应力下稍有改善，而 NMOS 晶体管则受到消极影响[10,11]。
- 即使更小尺寸晶体管采用 STI，ΔW 也变得重要。
- 现在漏电容的减小慢于面积的减小。
- 对小几何尺寸器件，掺杂损失和掺杂的统计波动增加了器件的波动性：输入/输出、模拟和存储器设计对此特别敏感。
- 为了控制漏感应势垒降低（Drain-induced Barrier Lowering，DIBL）效应而增加的沟道掺杂浓度则减小了载流子迁移率，而增加了体效应。
- 薄栅氧导致杂质穿透，将影响 PMOS 的驱动电流[6]。
- 随着栅氧厚度接近单层二氧化硅厚度，栅氧的缩小也变得越来越慢了（见图 1-1）。

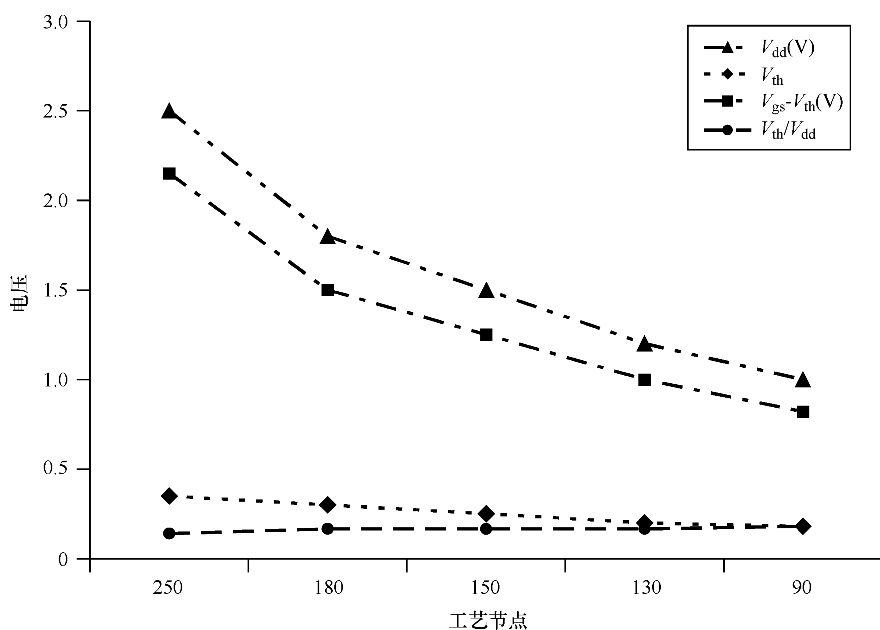


图 1-11　栅驱动与工艺节点的关系

泄漏问题　如果不能通过设计方法学的改进减缓其增加的趋势，亚阈区泄漏的增加速率最终会与芯片动态功耗（见图 1-12）的速率相同，特别是对几代工艺中的高性能微处理器而言。氧化层厚度每减小 1Å，栅电流（见图 1-13）会增

加 2.5 倍，从 130nm 开始，每代技术间的栅电流变化约为两个数量级。

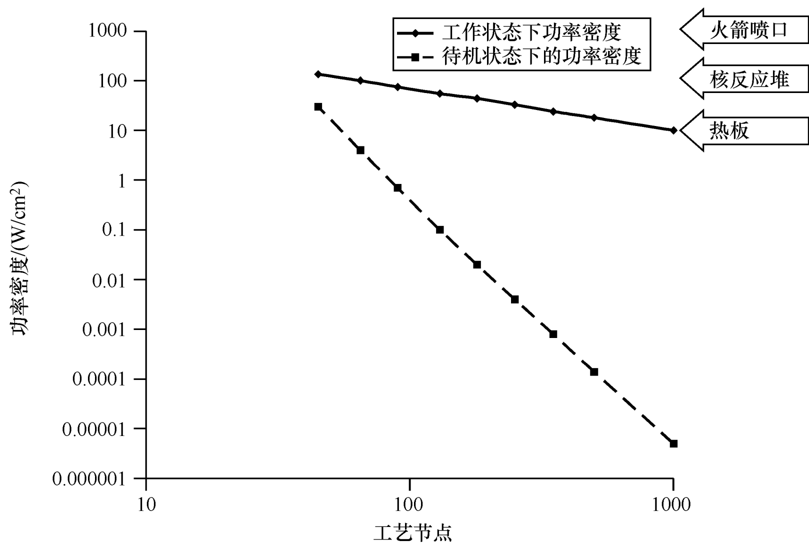


图 1-12　CMOS 功率密度的变化趋势

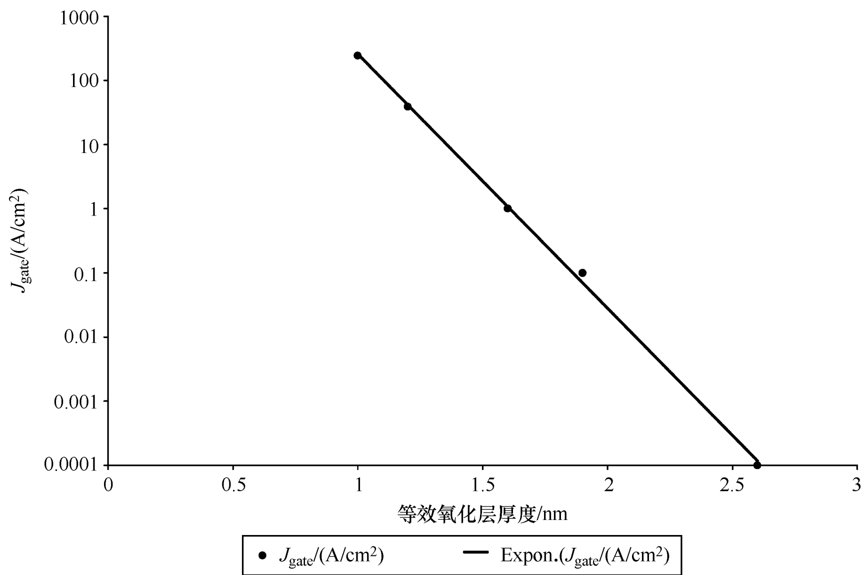


图 1-13　NMOS 的 J_{gate}（A/cm^2）与等效氧化层厚度的关系（数据源于 NEC）

随着特征尺寸的减小，栅电阻不断增加，并且 SD 电阻也随着结的不断减薄而增加。对结泄漏来说，需要仔细地调整 SD 电阻，直到需要的源漏结深是采用实际工艺可以实现的。图 1-4 表明，从 250nm 减小到 65nm，多晶栅厚度的变化

很小，发生显著变化的只是多晶栅的长度和宽度。因此，沟道电阻随着工艺缩小而增加并且需要在晶体管模型中予以考虑。

为了继续改善晶体管 $I_{d\ sat}$，人们对应变硅沟道晶体管投入了大量的认真研究和开发努力，据报道，采用 SiGe 应变硅已经实现了 10% ~20% 的改善[22]。一种初步的应变技术是采用氮化硅“帽”薄膜来提供应变沟道，只能实现适度的 $I_{d\ sat}$ 改善。提升源漏技术也在不断发展，但是该技术要求采用选择性外延工艺，这是一种困难的制造工艺。引入的许多新材料使高 k 栅氧化层成为现实，现在正采用 NiSi 取代 CoSi[22]。尽管高 k 栅氧化层可以与金属栅一起使用，但是与硅不兼容，因此在集成方面提出了重大的挑战[4]。与多晶栅相比，金属栅具有的明显优点是因为它们不会发生耗尽，所以对相同的电容有效氧化层厚度（CET），工艺工程师不需要使用更薄的栅氧化层[4]。因此，对给定的氧化层 CET，从理论上说金属栅技术的累积态栅泄漏更低。但是金属栅是非自对准的，因此为了采用金属栅，生产中必须进行革新。在准备本书时，预先掺杂的多晶硅正被用来减小多晶的耗尽问题，其代价是增加了刻蚀的复杂性。由于使用预先掺杂的多晶硅，一些厂商已经有了解决这些问题的方法。在集成电路生产的开始阶段，采用的材料只有 5 种，但是现在已经上升到 20 种了[23]。

由工艺技术缩小获得的性能改善已接近极限，但是尺寸的缩小有望会继续依据摩尔定律进行。现在，通过诸如新的晶体管设计和新材料及工艺改进可以改善性能，就最近开发的而言，包括高 k 介质、FinFET、SOI、应变硅以及纯硅衬底等。

1.4　工艺控制和可靠性

随着工艺缩小，栅长特征尺寸（CD）的绝对物理变化没有减小，因此对未来工艺节点来说，栅长 CD 变化的相对百分比会增大[7]。除此之外，栅长减小到 100nm 以下时，互连边缘的粗糙度（LER）影响着几个晶体管参数，因此受到了越来越多的关注。sub-100nm 工艺中 LER 控制是关键，因为随着工艺的缩小，对于更短的栅长，LER 对器件的影响来说非常重要。LER 控制需要采用先进的光刻和刻蚀工艺，这只能通过更好的工艺控制来改善。较大 LER 带来的负面作用是更大的交叠电容 C_{gd}，特别是对 PMOS。受到影响的其他器件参数包括 DIBL 的和阈值电压，因为退火后，有效沟道长度会随着 LER 减小，特别是 PMOS（见图 1-14）。当晶体管的 $L_{effective}$ 值由于 LER 效应而减小时，PMOS 的 V_{th} 和穿通电压将会受到相反的影响。

V_{th} 的变化受到掺杂的随机波动和栅 CD 变化的影响。薄栅氧与掺杂沟道的形成一起引起沟道杂质的波动，波动情况还与栅多晶的形态有关（参见图 11-7）。

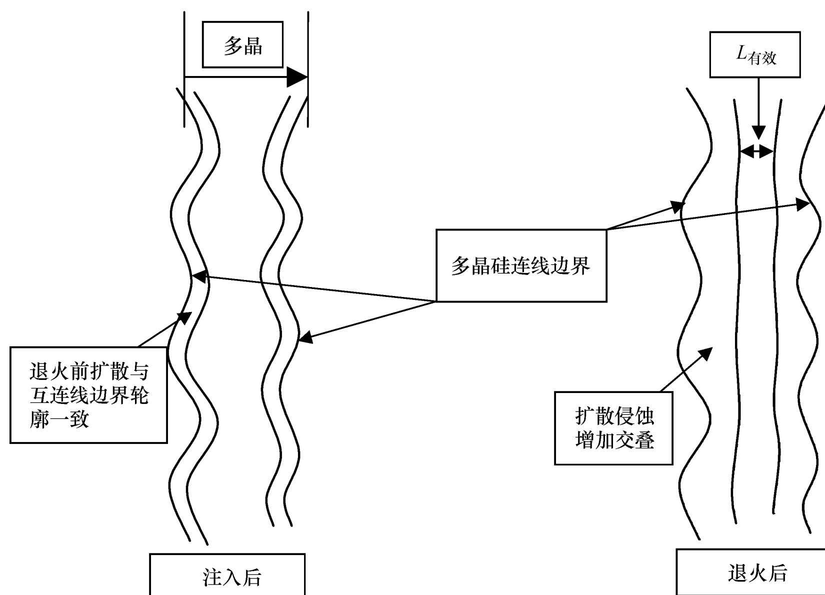


图 1-14　LER 使交叠电容增加而使沟道长度减小

这些效应使 V_{th} 更加难以控制，特别是晶体管 V_{th} 匹配更加困难，尤其是对小尺寸器件。从图 11-37 可以看到，最小器件 V_{th} 的变化最大，但其会逼近一个渐进值。避免使用最小宽度的晶体管是一个明智的选择，除非 V_{th} 变化不会引起电路失效。

负偏压温度不稳定性（Negative Bias Temperature Instability，NBTI）是栅氧厚度缩小后面临的一个效应。130nm 工艺节点的栅氧厚度已经对 NBTI 非常敏感了[18]。任何引起断键的工艺步骤都会增强 NBTI。特别是等离子或反应离子刻蚀是能引起键断裂的工艺，因此会使 NBTI 增强。65nm 工艺节点中，栅氧厚度将等于或低于 10Å㊀。对这个量级的厚度，界面控制非常关键。多晶耗尽将会是制约性能进一步缩小的限制因素，为此需要采用无耗尽的栅材料。90nm 及以下工艺节点的栅氧厚度控制对保持一个可预期的、较小的栅电流来说是很重要的。栅氧厚度每减小 1Å，栅电流大约增加 2.5 倍（见图 1-13）。

1.5　光刻问题和掩膜数据爆炸

从 180nm 节点起，开始进入亚波长范围，即最细线条已小于光刻中曝光采用的光源的波长。采用光学方法进行光刻越来越不能满足亚波长光刻的需求

㊀　1Å = 0.1nm，后同。

（见图 1-5），因为必须克服众多的困难才能使新一代光刻技术用于生产中。因此在 sub-100nm 节点，必须变革物理设计，从而不需要下一代光刻技术就可以进行可靠的图形光刻。在 90nm 节点以下，将需要采用先进的 OPC 技术，同时必须采用能够与光刻技术相适应的物理设计。分辨率扩展技术导致了图形分割之后的掩膜数据剧增，这也就增加了掩膜版费用[8]。不断变大的亚波长差距使得未来几代工艺节点中掩膜版和光刻费用成指数增加；因此，只有资金雄厚的生产厂商才能使用得起处于领先地位的光刻设备。对于其他厂商来说，在增加光刻设备数值孔径值和采用积极的 OPC 技术以扩展 193nm 光刻技术分辨率的同时，物理设计的自由度将不得不受到限制[9]。关于这个问题的详细介绍见第 3 章。

1.6　新型的电路和物理设计工程师

目前 CMOS 工艺的缩小已到达这样一个阶段，即从工艺开发的角度考虑，认为电路和物理设计之间相互没有关联的这种传统假设已不再成立。因此，这要求以我们在电路的实现方式方面发生模式的转换[20]。即使是可以保证性能的专用集成电路（Application-Specific Intergrated Circuit，ASIC）设计，如果要求能实现功能并且可以缩小到图形特征尺寸在 100nm 以下，设计方法也必须采用这种转换。

特别是高性能设计会要求完全不同的方法。这就要求新型的电路和物理设计工程师。他们清楚这些困难，通过建立与光刻技术相适应的物理设计，从而获得一个鲁棒的、可缩小的并且是高成品率的设计，使其成为解决方案的一部分。这种设计在未来的工艺节点中必须可以容忍较大的泄漏，包括亚阈区泄漏（包括 GIDL）和栅泄漏。波动容限是未来工艺节点中的另一个设计要求。

许多工艺步骤受版图设计风格的影响。最显著的是，多边形图形密度对层间介质层厚度有重要的影响。扩散层图形密度对最终产品制造成品率的影响很大。在器件匹配性对电路性能非常重要的电路中，其他版图类型可以缓解掺杂波动和多晶 CD 的变化。

新型的电路和物理设计工程师必须清楚邻近效应对电路以及随之对电路设计的相应影响，从而使硅集成电路能够如仿真时预测的一样。如果将一个晶体管紧靠着阱或位于多晶密集或稀疏的区域时，都会产生邻近效应。如果将晶体管紧靠另一种结构，注入时将使注入的杂质偏转到紧靠光刻胶掩膜的晶体管上，引起掺杂波动。只要每个晶体管的相邻环境都相似，邻近效应就是一致的。否则邻近效应会引起器件 V_{th} 的变化。其他的邻近效应包括光刻引起的多晶 CD 变化、刻蚀微负载导致的欠优化版图风格引起的刻蚀邻近效应，以及光学邻近效应。许多系统的邻近效应可以通过采用好的版图风格和依靠光刻技术以及偏置加以避免。但

是设计者必须理解邻近效应带来的制约因素，并且能够采用设计技术来减轻这些效应的影响。本书将在后面章节中详细地讨论这些技术，以便给电路和物理设计工程师提供一个知识背景，使他们可以通过物理设计更好地应对这些效应的影响。

1.7　建模的挑战

持续的物理尺寸缩小增加了器件电参数的容差，并且成为对建模的挑战。在 BSIM4 之前，模拟软件中没有包含栅电流模型，设计人员不得不自己考虑栅电流的计算。沟道中杂质的统计波动将影响位单元中使用的小几何尺寸晶体管，但是对其建模难度较大，使得建立的模型精度较差[3]。由于邻近效应和 STI 应力导致的迁移率退化与版图本身的关系非常密切，因此建模难度也很大[10,11]。目前已经推出了一些新的工具，通过版图提取在这个领域提供一些帮助。最好的工作方式是理解这些效应，然后在物理设计中使这些效应对电路的影响最小。第 2 章将详细讨论这些效应。

由于 halo 注入（大斜角注入）效应与沟道长度之间关系的无规律性，导致对 halo 注入的逻辑过程进行的模拟建模精度很差。这样混合信号设计工程师就需要使用“模拟晶体管”[17]，这将增加费用，并且有时还没有模拟晶体管可供使用。除非你能够与 Foundry 生产线共同工作，有能力将 halo 对 DIBL、V_{th}以及厄尔利电压的影响及其和晶体管沟道长度的关系进行建模，否则明智的做法还是采用模拟晶体管。正如 IEDM 2002 出版的参考文献［16］中所述，这样一种模型并不是不可能的，但是并不是每条 Foundry 生产线都能提供这种模型。无论什么原因，如果你必须在模拟设计中使用 halo 工艺的晶体管，而 SPICE 模型又没有考虑 halo 效应（反型短沟道和漏感应阈值电压漂移）和输出电阻以及厄尔利电压的变化，选择合适的晶体管尺寸就变得非常重要。要求模型能够适用于选用的晶体管尺寸，以避免由于晶体管特性随着沟道长度的非线性变化而影响模拟精度。

需要建立模型的新物理效应包括对 V_{th}与晶体管多晶长度之间关系产生影响的 halo 注入效应（反型短沟道（RSC）效应）[16]、栅感应漏极泄漏（GIDL）、漏感应阈值电压漂移（DITS）[24]、输出电阻和厄尔利电压变化，以及栅电流[15]。在以往的模型中，仅从 BSIM4 开始才对其中一部分新效应建模[24]。对小于 130nm 的工艺节点，强烈推荐在所有的仿真中，包括在数字电路仿真中，采用 BSIM4 模型。

针对上面描述的注入工艺和特征尺寸的波动，统计模型对解决其中一些问题也是必需的。除非能明智地选择适合特定电路的模型组合，否则只是采用工艺角模型会以速度、功耗和面积为代价，产生不切实际的工艺组合，导致对电路过于保守的设计。另外，通过传统的五点工艺角方法并不能建立关键工艺角模型，因此对特定的电路来说，可能并没有分析最坏情况。表 1-1 是对建模挑战的总结。

表 1-1　sub-100nm 工艺中器件建模挑战的总结

参　　数	效应产生原因	效 应 说 明
RSC	halo 注入（工艺、物理器件效应）	由于横向不均匀掺杂引起的反型短沟道效应；当沟道长度变化时，V_{th}随之变化
DITS	halo 注入（工艺、物理器件效应）	对沟道较长的器件，当 halo 注入对沟道的影响减小时，由于 DIBL 的变化导致的漏感应随阈值电压漂移
厄尔利电压和输出电阻[17]	halo 注入（工艺、物理器件效应）	对长沟器件，DIBL 变化导致的效应，与上面相似
多晶耗尽[25]	超薄栅氧（工艺、物理器件效应）	对超薄栅氧，多晶耗尽正变得很重要，对大多数器件来说多晶耗尽大约导致 8nm 等效氧化层厚度的增加（EOT），对预掺杂的多晶则很小
栅隧穿电流	超薄栅氧（工艺、物理器件效应）	由于超薄栅氧，从栅到沟道发生直接遂穿
迁移率随掺杂的变化	halo 注入（工艺、物理器件效应）	迁移率随着掺杂的降低而改善
线性邻近效应	密集、隔离	部分由光刻效应引起，部分与刻蚀微负载效应有关，还与多晶对杂质的散射有关，导致掺杂与设计中多晶互连间距有关的系统变化
非线性邻近效应	光学邻近修正（OPC）	亚波长光刻技术要求分辨率增强技术
GIDL	能带到能带之间的隧穿	对于高的结掺杂和 sub-100nm 器件中的突变结，由漏到栅之间的强电场引起能带到能带的隧穿
扩散和多晶光刻后直角弧化	工艺和版图效应	亚波长光刻引起扩散和多晶光刻后直角弧化，进而导致小尺寸器件的器件图形变化以及多晶接触压焊点到扩散边界的邻近效应
阱邻近	在阱边界上的器件	光刻胶外的阱注入原子的横向散射会导致靠近阱边缘的器件阈值电压的增加。对 NMOS 和 PMOS 器件，典型增加量分别为 50mV 和 20mV
STI 应力	STI 对器件沟道的邻近效应	STI 应力减小电子迁移率但是增加空穴迁移率，因此影响 $I_{d\ sat}$

1.8　变革设计方法的需要

过去，进行容性噪声分析就足够了，但是现今，信号完整性已经扩展到了感

性噪声。时序通常是过去主要关心的问题，现在也要考虑功能问题了。为此就需要开发抗噪声电路以减小长时分析以及对片上和片外信号完整性的建模。同时还需要开发通过构建进行校正的技术以保证信号完整性的信号传输方法。这可能会是采取放置中继放大器的方式，在间距允许的地方展开互连。在某些地方，可能需要屏蔽。对感性屏蔽，鲁棒的功率分配系统将翻倍，还要求采用宽大总线返回路径。

最近，伴随着外加电压缩小、器件尺寸缩小以及更高的时钟频率，也需要考虑功率完整性问题。尽管器件尺寸不断缩小，但是为了满足不断增加的芯片性能的需求而使功能不断增加，导致功率耗散始终保持向上增加的趋势。当电源电压下降时，功耗却在增加，因此电源电流不断增加并且因此 $\mathrm{d}i/\mathrm{d}t$ 和阻性压降也不断增加，使得 $L(\mathrm{d}i/\mathrm{d}t)$ 正逐渐成为主要的性能制约因素。为了应对这个问题，现在设计方法必须能够将芯片的功率分配设计扩展到封装和系统板以实现整个系统的解决方案。否则，将不可能实现期望的电源阻抗，从而缓解较高的阻性和 $L(\mathrm{d}i/\mathrm{d}t)$ 压降。

工艺的波动，不管是器件还是互连的波动都将是纳米 CMOS 设计的主要问题。对于能适应较大波动的设计，设计方法必须具有应对波动的能力。传统的五点工艺角方法正变得越来越没有意义。在某些情况下会以芯片面积和功耗为代价导致一个开销巨大的过于保守的设计，而在其他情况下又可能完全错失了重要的最坏情况条件。

设计方法的自由度正在减小。未来的设计将会提出对准关键多晶的要求。这也要求在位单元设计中进行变革。现有的设计中传输晶体管的多晶与下拉和上拉晶体管直交。新的位单元设计需要考虑这个问题，并且使所有的多晶互连对齐。使所有的多晶互连以同一方向排列的原因是由于采用了有一定注入角度的 halo 注入。如果将栅布置成彼此之间直交，会导致掺杂的变化，这是由于每个多晶栅边缘接受 halo 注入的时间不同而引起的。例如，水平方向栅在不同时间只接受到一半的注入剂量，因此引起 V_{th} 的波动。由于光刻效应和掩膜版的影响，彼此直交的多晶互连的 CD 变化也更高，详细情况在第 11 章介绍。

泄漏（亚阈区、GIDL 和栅）是我们在新的设计方法中必须面对的下一个难题。存储器必须设计成比以前能容忍更大的泄漏，但是又不能明显降低阵列的效率。在像 L2 和 L3 缓存那样巨大的阵列中，较大的泄漏不仅是性能和功能问题，也是面积和功耗问题。将 L2 和 L3 缓存设计成采用多个周期中存取可能是必需的，因为它们可以容忍更长的等待时间。这是弥补更慢的存取时间所必须的，因为采用更长的沟道长度和更高的 V_{th} 注入来减小泄漏功耗是以增加存取时间为代价的。由于较长的沟长可以使位单元晶体管实现更好的匹配，并且允许采用更富挑战的下拉/传输晶体管，这也可使速度得到一定的恢复。

宽的多米诺门不再是纳米 CMOS 时代切实可行的设计类型，这是由于在功

能、噪声容限和速度之间进行折衷的难度更大了。具有功能的宽多米诺电路将不再比一个采用两级电路的单元快了。比例逻辑也将被摒弃。器件和泄漏波动将导致设计良好的比例逻辑偏离其最佳工作点，某些情况下甚至会完全失去功能。

功率耗散、性能和工艺复杂度之间协调变得更加困难，为此要求设计者更加慎重地确定对不同晶体管可以采用的阈值电压注入的最佳次数，并且与费用之间进行权衡。如果设计中能够采用较低 V_{th} 的晶体管，设计人员就可以在改善芯片性能的同时不会大幅增加待机功耗。

在 130nm 工艺中，互连线转换到采用铜互连已经表明能显著提升抗电迁移（EM）能力和互连性能。然而，当芯片面积增大时，设计者要求更高的互连性能。而在 130nm 后的各代工艺中还没有出现互连线的进一步更新。随着工艺尺寸的缩小，工艺工程师正试着采用低 k 介质提升互连性能。因为低 k 介质的热传导性更差，在纳米 CMOS 工艺中，EM 问题已经重新成为一个问题。加之更高信号速度导致流过互连的电流脉冲更高，进一步恶化了 EM 问题。

1.9　总结

本章讨论了特征尺寸缩小到 100nm 以下时所带来的大部分问题，并指出，如果我们继续使用先前工艺中开发的设计方法，这些问题将提出哪些挑战。很清楚的一点是我们需要进行模式转换，才能在未来的设计中继续利用工艺技术缩小的优势，从而继续沿着摩尔定律确定的趋势发展[27]。

虽然我们看到作为器件尺寸缩小的结果，性能改善的趋势已经变缓。但是如果器件和工艺工程师发明了新的工艺和材料，能够解决由于物理限制影响性能改善的问题，性能的进一步改善是可以持续的[23]。

虽然如此，现在已经到达这样一个阶段，即要求电路和物理设计工程师也必须理解尺寸进一步缩小带来的这些效应，以便利用这种技术并且确保具备功能的鲁棒设计。随着掩膜费用的增加，更迫切的问题是工程师们能理解这些效应，以避免陷阱，并在第一次投片中就能实现设计功能。

参 考 文 献

[1] *IBM J. Res. Dev.*, Vol. 46, No. 2/3, 2002.

[2] P. Kapur, Performance challenges of the future on chip metal interconnects and possible alternatives, Stanford University, May 23, 2002.

[3] Near limit scaling, workshop, Solid State Circuits Technology Committee, 2003.

[4] The future of semiconductor manufacturing, short course, *IEEE International Electron Devices Meeting*, 2002.

[5] S. Schulze, Mentor Graphics Corp., Wilsonville, OR, Effecting mask costs by solving the data explosion bottleneck in mask data preparation, *Semiconduct. Int.*, July 1, 2003.

[6] H. S. Momose, S. Nakamura, T. Ohguro, T. Yoshitomi, E. Morifuji, T. Morimoto, Y. Katsumata, and H. Iwai, Study of the manufacturing feasibility of 1.5 nm direct-tunnelling gate oxide MOSFETs: uniformity, reliability, and dopant penetration of the gate-oxide, *IEEE Trans. Electron Devices*, Vol. 45, No. 3, Mar. 1998.

[7] A. Allan, D. Edenfeld, W. Joyner, A. Kahng, M. Rodgers, and Y. Zorian, International Technology Roadmap for Semiconductors, *IEEE Comput.*, Jan. 2002.

[8] S. Schulze, Effecting mask cost by solving the data explosion bottleneck in mask data preparation, *Semiconductor Int.,* July 1, 2003.

[9] Y. Pati, Sub-wavelength lithography, Tutorial, *Design Automation Conference*, 1999.

[10] C. Diaz, M. Chang, T. Ong, and J. Sun, Process and circuit design interlock for application-dependent scaling tradeoffs and optimization in the SoC era, *IEEE J. Solid State Circuits*, Vol. 38, No. 3, Mar. 2003.

[11] G. Scott, J. Lutze, M. Rubin, F. Nouri, and M. Manley, NMOS drive current reduction caused by transistor layout and trench isolation induced stress, *IEEE International Electron Devices Meeting*, 1999.

[12] M. Horowitz, R. Ho, and K. Mai, The future of wires, *Semiconductor Research Corporation Workshop on Interconnects for Systems on a Chip*, May 1999.

[13] V. Agarwal, M. Hrishikesh, S. Keckler, and D. Burger, Clock rate vs. IPC: the end of the road for conventional microarchitectures, *27th Annual International Symposium on Computer Architecture*, June 2000.

[14] T. Sakurai, Issues of current LSI technology and an expectation for new system-level integration, *International Conference on Solid State Devices and Materials*, pp. 36–37, Sept. 2001.

[15] K. Osada, Y. Saitoh, E. Ibe, and K. Ishibashi, 16.7fA cell tunnel-leakage-suppressed 16 Mb SRAM for handling cosmic-ray-induced multi-errors, Session 17.2, *International Solid-State Conference*, 2003.

[16] R. Rios, W. K. Shih, A. Shah, S. Mudanai, P. Packan, T. Sandford, and K. Mistry, A three-transistor threshold voltage model for halo processes, *IEEE International Electron Devices Meeting*, Dec. 2002.

[17] K. Cao, W. Liu, X. Jin, K. Vasanth, K. Green, J. Krick, T. Vrotsos, and C. Hu, Modeling of pocket implanted MOSFETs for anomalous analog behavior, *IEEE International Electron Devices Meeting*, 1999.

[18] C. Liu, M. Lee, C. Lin, J. Chen, Y. Loh, F. Liou, K. Schruefer, A. Katsetos, Z. Yang, N. Rovedo, T. Hook, C. Wann, and T. Chen, Mechanism of threshold voltage shift (ΔV_{th}) caused by negative bias temperature (NBTI) instability in deep sub-micron pMOSFETs, *Jpn. J. Appl. Phys.*, Vol. 41, Pt. 1, No. 4B, pp. 2424–2425, Apr. 2002.

[19] A. Stamper, Interconnection scaling to 1 GHz and beyond, *MicroNews*, Vol. 4, No. 2, first quarter 1998.

[20] International Technology Roadmap for Semiconductors, *http://public.itrs.net.*

[21] P. Ranade, H. Takeuchi, W. Lee, V. Subramanian, and T. King, Application of silicon–germanium in the fabrication of ultra-shallow extension junctions for sub-100 nm PMOSFTs, *IEEE Trans. Electron Devices*, Vol. 49, No. 8, Aug. 2002.

[22] S. Thompson et al., A 90 nm logic technology featuring 50 nm strained silicon channel transistors, 7 layers of Cu interconnects, low κ ILD, and 1 μm^2 SRAM cell, *IEEE International Electron Devices Meeting*, 2002.

[23] A. Grove, Changing vectors of Moore's law, *IEEE International Electron Devices Meeting*, 2002.

[24] J. Assenmacher, *BSIM4 Modeling and Parameter Extraction*, CL TD SIM, Infineon Technologies, Workshop Analog Integrated Circuits, Berlin, Germany, Mar. 19, 2003.

[25] C. Choi, Modeling of nanoscale MOSFETs, Ph.D. dissertation, Stanford University, 2002.

[26] G. Brown, The tyranny of roadmap: new CMOS gate dielectrics with reliability promises and challenges, ISMT Reliability Engineering Working Group, Dec. 12, 2001.

[27] G. Moore, Cramming more components onto integrated circuits, *Electronics*, Vol. 38, No. 8, Apr. 19, 1965.

[28] G. Moore, No exponential is forever..., keynote, *IEEE International Solid-State Circuits Conference*, 2003.

第 2 章　CMOS 器件与工艺技术

2.1　前道工序的设备要求

过去 10 年，集成电路技术领域取得了重大的突破。在集成电路生产的后道工序，由于采用特有的集成方式（如双镶嵌工艺），在互连技术引入了铜互连及低 k 介质，使 RC 得到了改善。前道工序中，高性能器件的栅介质通常采用由单层原子层淀积生长形成的可靠的氮氧化合物。伴随着超浅结技术的重要进展，应变工程使得采用 sub-130nm 技术生产高功率和低功率器件成为可能。本章我们将回顾 sub-130nm 技术前道工序生产过程目前的状态及未来可能的发展趋势。

2.1.1　技术背景

在过去的 40 年，以快速步伐不断前进的半导体工业提供了速度更高、功能更强大、更小型化、价格更低的电子产品。由于 MOSFET（金属氧化物半导体场效应晶体管）是集成电路中使用最普遍的单元，因此半导体工业生产中缩小 MOSFET 尺寸的能力更是极大地推动了电子产品市场的快速发展。尽管存在很多挑战，但是摩尔定律依然在 IC 的晶体管尺寸缩小方面起着指导性作用。长期以来，栅长和栅氧厚度的缩小一直是影响晶体管尺寸缩小的两个关键因素。栅长小于 130nm 的生产技术以及栅长小于 30nm 的技术研究都分别取得了很大的进展。然而，超薄栅氧中基本量子力学的隧穿效应以及对短沟道效应和关态电流的控制程度，限制了由尺寸缩小带来的预期性能改善。因此，在传统的 MOS 架构中采用了新的维数。例如通过多种技术（如采用 SiGe）提高沟道迁移率的应变工程。另外一种方式就是摒弃体平面晶体管结构，采用绝缘上硅（SOI）和三维 finFET（鳍形栅 FET）器件。

基于硅的集成电路典型工序首先是生成隔离结构。易失性（如 SRAM）和非易失性［如 flash，见图 2-1a］器件工艺中采用浅槽和深槽。槽的刻蚀是一个很大的挑战（例如如何保证获得正确的侧墙剖面）。槽的填充也极其关键。介质材料的合理选择是减少硅器件沟道中空洞、寄生结和不希望产生的应力的关键。槽填充后，一般用电机械抛光的方法去除多余的介质材料。然后经过各种湿法及干法清洗工序清洗硅表面，为后面的离子注入及栅介质淀积做准备。在多晶硅刻蚀后，变化的低能量倾斜角度的离子注入、侧墙偏移和短时快速退火等工序生成超

浅结。然后形成低热消耗隔离层以及随之的源漏注入和激活，再接着是形成低热消耗的金属硅化物。采用氮化层作接触孔刻蚀阻挡，允许位于槽氧化层上的接触与源漏接触有一定量的偏移。最后，一层淀积薄膜如硼磷硅玻璃（BPSG）或高密度等离子（HDP）薄膜形成第一层层间介质层，到此完成了生产的前道工序。闪存生产过程的工序与此相似。采用约 10nm 的隧道氧化层，使得沟道热电子注入到多晶硅Ⅰ，即浮栅中，随后利用 Fowler-Nordheim 隧穿效应擦除单元。闪存单元中有时使用不对称源漏结构，还采用氧化层-氮化物-氧化层的层叠结构形成浮栅多晶硅Ⅰ和控制栅多晶硅Ⅱ之间的第二层栅介质[2,3]。本章将回顾前道工序和在 sub-130nm 工艺中可使用的关键设备，重点是栅层叠结构、应变工程和快速热处理工艺。

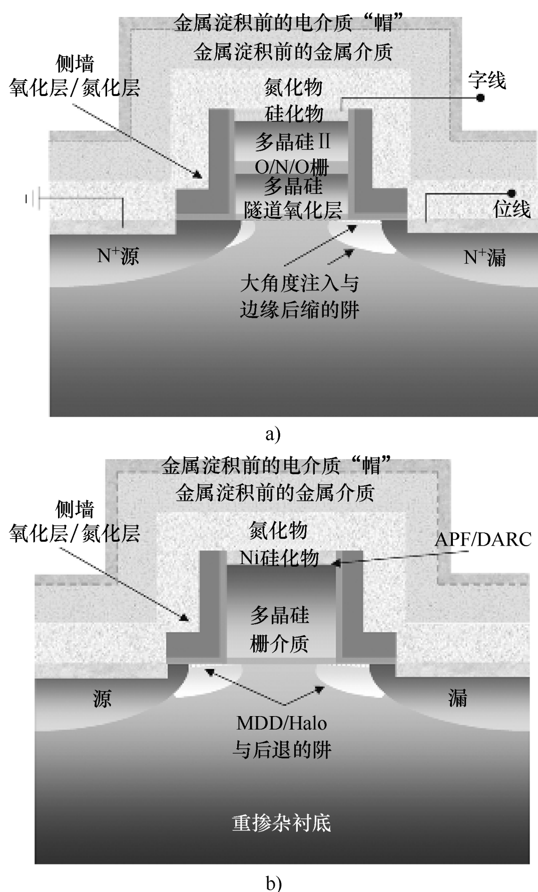


图 2-1　a）典型的 NMOS 闪存单元和 b）典型的 MOSFET 单元

［虽然这两种单元间大多数的工艺步骤都是相同的，但是每种单元的设计和工艺还是要面对一些独特的挑战。例如，MOSFET 的栅缩小到了 sub-2nm，而闪存单元的栅（由于其对栅泄漏电流的要求）却在 sub-10nm 范围］

2.1.2 栅介质的缩小

随着 MOSFET 尺寸的缩小（见图2-1b），要求提高单位面积栅介质电容，因此就要减小栅介质厚度。本节将回顾半导体工业可能选用的新栅介质代替 SiO_2 时所面临的挑战与栅介质缩小的趋势[1]。栅层叠结构由栅介质（SiO_2 或者 SiON）和其上高掺杂的 N^+（NMOS）和 P^+（PMOS）多晶硅栅电极组成。为了改善性能、增加密度和更好地控制短沟道效应，器件尺寸的缩小要求栅介质也随之减薄。当栅氧厚度缩减到 4.0nm 时，工业生产面临着新的挑战。这些挑战包括：PMOS 高掺杂多晶硅电极中硼的穿通、增加的泄漏和日益严重的可靠性问题。生产中，采用由热氮化过程（N_2O、NH_3 或者 NO）生长的氮氧化硅（SiON）来阻止硼穿通氧化层并提高其抗热载流子的能力。当栅介质厚度减小到 2.0nm 时，采用等离子氮化硅作为栅介质，以提高介质中的氮分，并较好地控制氮的分布[4]。尽管人们曾经担心超薄氧化层的击穿将成为主要的可靠性问题，但是由于电源电压的不断减小，使超薄氧化层可以继续满足可靠性要求。然而，由于穿过超薄氧化层的栅泄漏电流将产生过大的待机功耗并使介质的完整性和可靠性退化，这可能成为介质进一步缩小的制约因素。

栅介质中的载流子输运 SiO_2 高达 9eV 的禁带宽度和巨大的势垒高度使硅氧化层在中度偏置条件及厚度大于 4.0nm 时接近于理想的绝缘体。这与其他一些薄膜（例如 Si_3N_4 或者高 k 介质）形成对比。在 Si_3N_4 或者高 k 介质薄膜中，传导特性由体制约机制（如 Frenkel-Poole 发射）表征[5,6]。将电子由费米能级激发到真空需要的能量称为功函数 ϕ_m。在外加偏置 $V_{ox}=E_{ox}t_{ox}$ 作用下，电子有一定的概率隧穿 Si-SiO_2 势垒，由 Si 的导带到达 SiO_2 的导带。穿过三角势垒的传导可由 Fowler-Nordheim 隧穿来描述，电流密度可表示为[7-9]

$$J_{FN}=AE_{ox}{}^{2}\exp\left(\frac{-B}{E_{ox}}\right) \qquad (2.1.1)$$

式中，A 是与 Si-SiO_2 势垒高度 ϕ_b 相关的常数；B 是与电子有效质量 m^* 和 ϕ_b 有关的常数。当氧化层厚度减小并且 V_{ox} 下降时，电子不再进入导带而直接隧穿梯形势垒。V_{ox} 小于势垒高度 ϕ_b 时，直接隧穿电流密度可由式（2.1.2）来描述[10,11]。

当介质厚度小于 3.0nm 时，直接隧穿电流将是主要的电流传导机制。由于隧穿电流与氧化层的厚度成指数关系，介质厚度缩小到 1.0nm 时，将引起不希望的高泄漏电流，导致高待机功耗、可靠性问题及介质完整性问题。NMOS 泄漏电流将是栅介质缩小的制约因素。由于空穴隧穿要通过较高的势垒，因此 NMOS 隧穿电流约是 PMOS 隧穿电流的 10 倍[11]。

$$J_n=AC(V_g,V_{ox},t_{ox},\phi_b)\exp\left\{\frac{-B[1-(1-V_{ox}/\phi_b)^{3/2}]}{E_{ox}}\right\} \qquad (2.1.2)$$

式中，$C(V_g, V_{ox}, t_{ox}, \phi_b)$ 是修正函数，与 V_g、V_{ox}、t_{ox}、ϕ_b 有关，由经验拟合获得[11]。

***C-V* 和等效氧化层厚度**　低频和高频的电容电压（*C-V*）特性测量通常可以用于提取金属-绝缘层-半导体（MIS）特性，包括介质厚度、平带电压、固定电荷和界面态密度等。然而对薄栅氧，特别是当栅氧厚度小于 2.0nm 以下时，*C-V* 数据的测量和解读变得非常复杂。随着介质厚度的减小，穿过薄介质的隧穿电流成指数增长关系（氧化层物理厚度每减小 0.2nm，隧穿电流约增大 10 倍），在栅电极和衬底中存在的串联电阻上将产生压降（见图 2-2）。栅介质可以等效为一个与电压有关的电阻与电容的并联，栅电极和衬底起分布串联电阻的作用[12]。在强反型中，沟道阻抗引起的电容衰减将变得很重要，这就制约了测量 MOSFET 时可以采用的器件沟道长度[13]。

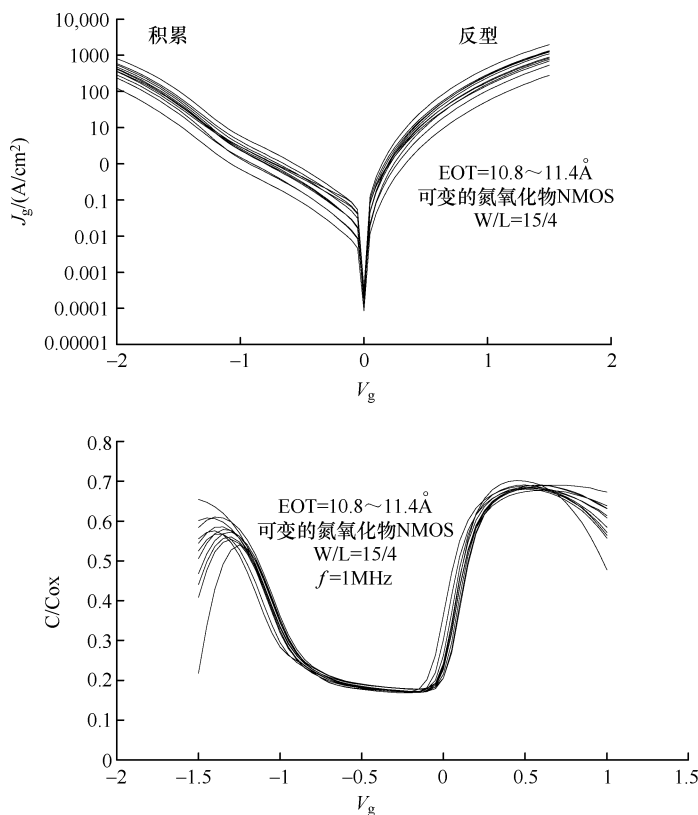


图 2-2　薄氮氧化物的 $J_g - V_g$ 即 *C-V* 曲线

（不断增加的氮分降低了隧穿泄漏，将减小电容在反型和积累时的衰减）

（数据由应用材料公司提供[13]）。

近年来，许多重要的工作都集中在精确的测量以及对电容电压曲线的解读上，这在本节的参考资料中有所体现。介质的电学厚度是栅中电荷的质心与衬底之间的距离[6]。在栅介质界面附近的多晶硅中，可动电荷载流子的耗尽，特别是反型时，将使电荷质心偏离界面约 0.3nm。这个效应可以建模为氧化层电容串联了一个附加电容[5]，这将使得测定的电学厚度大于预期的厚度。相似地，在衬底的反型层或积累层中，载流子被局限在一个靠近表面的很窄的势阱中，它们在表面法线方向上的运动要通过量子力学来分析。一个简单的解析表达式处理是不合适的，而正确的处理要求耦合有效质量的薛定锷方程和泊松方程的自洽求解[14]。用量子力学的观点分析反型层，发现反型电荷的质心偏离界面 0.3nm。由于多晶硅耗尽和量子力学效应导致的电学厚度在超薄介质中变得日益重要[5-15]。因而实际测定的电容值与预计值之间产生了巨大的差异。

电容的有效厚度（Capacitance Effective Thickness，CET）是指电容介质的电学厚度，可表述为[12]

$$\mathrm{CET}(V)=\frac{\varepsilon_0\varepsilon_{\mathrm{SiO_2}}A_{\mathrm{gate}}}{C(V)} \tag{2.1.3}$$

式中，ε_0是真空介电常数；$\varepsilon_{\mathrm{SiO_2}}$是 SiO_2 的介电常数；A_{gate}是栅面积。C（V）是与给定电压 V 对应的电容，它包括了由于多晶硅耗尽和衬底的量子力学效应所增加的串联电容。因此 CET 与电极的类型、电极的功函数以及电极的耗尽有关，也与衬底掺杂和栅电压有关[12]。

相比之下，介质的等效氧化层厚度（Equivalent Oxide Thickness，EOT）与电极的性质以及衬底的掺杂均无关。EOT 是指这样的等效氧化层厚度，由 EOT 厚度氧化层产生的 *C-V* 曲线与采用替代介质的 *C-V* 曲线相同，可以由下式得出[13]：

$$\mathrm{EOT}=\frac{\varepsilon_{\mathrm{SiO_2}}}{\varepsilon_{\mathrm{high\text{-}k}}}\cdot t_{\mathrm{high\text{-}k}} \tag{2.1.4}$$

式中，$t_{\mathrm{high\text{-}k}}$是高 k 介质的物理厚度；而 $\varepsilon_{\mathrm{high\text{-}k}}$是介质的介电常数。由于 SiON 或者其他中、高 k 介质的介质常数不知道，EOT 必须依照上面描述的电容测量方法确定来获得[12]。一旦 *CV* 测试完成，接下来的数据校正和解读将是一个挑战。人们提出多种不同模型来分析多晶硅耗尽及量子效应并进而提取 EOT。不同的模型和算法将导致提取的 EOT 结果不同，因此对比由不同方法获得的介质 EOT 时，必须多加注意[12,13,16,17]。

制约超薄 SiO_2 缩小的因素与替代介质　如前所述，栅介质缩小可以改善器件性并能抑制短沟道效应。但是有几个基本因素制约着 SiO_2 进一步缩小到小于 1.0nm EOT 的水平。当介质材料的禁带宽度达到 SiO_2 的禁带宽度时，每个界面的厚度大约是 0.35～0.4nm，这样两个界面总厚度就达到 0.7～0.8nm[6]，因此 SiO_2 缩小的绝对物理限制就是 0.7nm。然而随着氧化层厚度的减小，包括过多的

泄漏和有限的甚至零性能增益等在内的其他实际的限制可能会先达到。如式（2.1.2）所示，随着介质物理厚度的减小，隧穿电流将指数增加。另外，随着介质厚度的缩小，硅沟道和多晶硅电极的界面对 EOT 以及沟道迁移率的相对作用增大[5,6]。据报道，随着介质厚度的日益减小，较薄氧化层中较大的迁移率退化将导致 $I_{d\ sat}$下的增益比预期的小得多[62]。

硅的氮氧化物可以通过热氮化，或在 NO、N_2O 或 NH_3 中退火，或由 SiO_2 的等离子氮化（见图 2-3）等方式生成。介质中加入氮将通过几种途径改变材料的特性（见图 2-4）。硼的穿通将使 PMOS 的 V_{th}发生很大的漂移，并使介质的可靠性衰退，而氧化层中的氮对硼穿通则起到势垒阻挡作用。随着氧化层中氮分的增加，SiO_2 的折射率也将从 $\eta_{SiO_2}=1.46$ 增加到 $\eta_{Si_3N_4}=2.0$。另外，相对介电常数将随着氮的增加而线性增加，由 $k_{SiO_2}=3.9$ 增大到 $k_{Si_3N_4}=7.5$。如式（2.1.4）所示，为了保证一定的 EOT，如果使用的介质具有较高的 k 值，就可以采用物理厚度相对较厚的薄膜，这样就可以减小隧穿电流[5]。然而，SiO_2 中添加氮将减小禁带宽度，导致电子及空穴隧穿的势垒高度（ϕ_b）降低[5,8,20,21]。这意味着因较大的 SiON 物理厚度而减小的直接隧穿效应将部分地被较小的有效势垒高度所抵消[18,20,21]。氮氧化物一般在一氧化氮（NO）中生长或者退火。采用 NO 生成氮化氧化层的过程中，氮的结合受到制约，并且氮一般只堆积在界面上。对超薄氧化层，必须提高氮的百分比（5%～20%）以进一步降低泄漏和阻止硼穿通[4]。在生长 sub-1.5nm 氧化层时，采用等离子氮化可以更好地控制介质中氮的百分比和分布[4,24,27-29]。

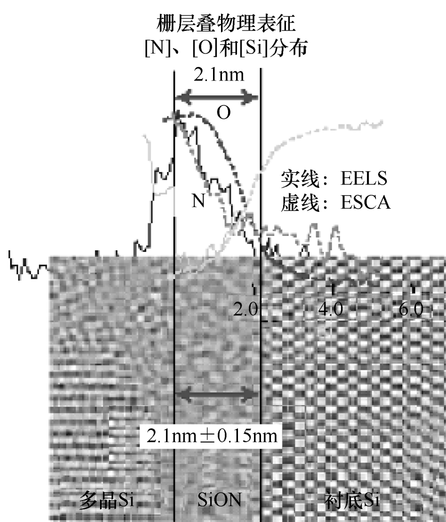


图 2-3　电子能量损失谱（EELS）和化学分析电子能谱（ESCA）

（测量表明：等离子氮化在多晶硅-氮氧化物界面引入有氮）（来源于参考文献［4］）

介质中的氮元素影响着 NMOS 与 PMOS 器件的迁移率。对 PMOS 器件，在所有的电场下，空穴迁移率随着氮分的增加而降低。而对 NMOS，在低氮分时，电子迁移率的峰值随着氮分的增加而衰退，但是高场下电子迁移率的下降随着氮浓度的增加而有所改善（见图 2-5）[23]。薄膜中大量的氮会在界面处产生陷阱或者成为载流子散射中心，使迁移率发生巨大的衰退[10]。氮对载流子迁移率的影响可以通过氮的剖面分布以及氮元素靠近沟道的程度来调制[22]。

为了抑制栅泄漏电流和持续实现介质缩小，针对有更高介电常数的替代介质

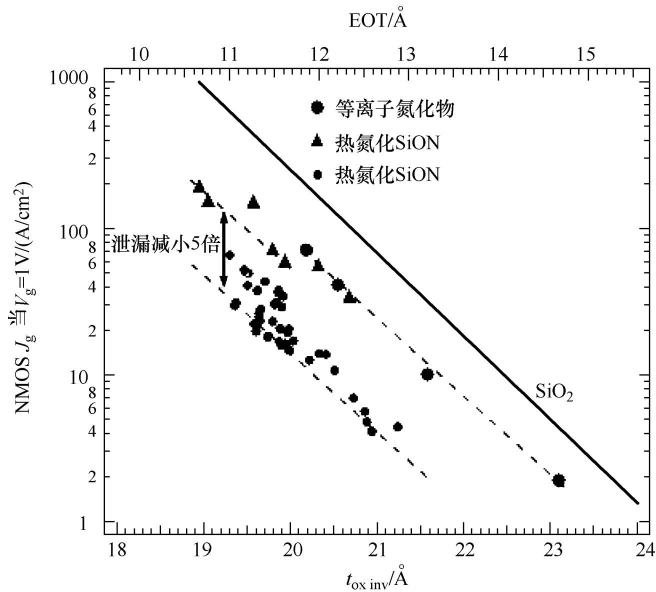


图 2-4 典型的热氮化和等离子 SiON 的 J_g-EOT 关系曲线

（如图 2-2 所示，增加氮含量使隧穿电流下降，数据由应用材料公司提供）

材料的热心研究一直在继续。具有高于 SiON 介电常数的材料，在产生同样的 EOT 时，采用的物理厚度可以扩大 $k/k_{Si_3N_4}$ 倍，因此根据式（2.1.2）将可以更好地抑制隧穿电流。氮化硅、氧化铝、氧化锆和氧化铪以及它们的硅酸盐都是正在研究的几种高 k 介质。替代介质的重要特性包括介电常数、禁带宽度、硅的能带对齐、热力学稳定性、界面质量、薄膜结构、可靠性、与栅电极以及 CMOS 工艺的兼容性等[5,6]。与高 k 栅介质材料相关的迁移率衰退（特别是采用金属栅时）问题的改善工作已经取得了重大的进展。目前，与 HfSiON 相关的极富意义的成果已有报道[25]。然而，其他可能的基本特性，如多晶硅金属氧化层界面的费米能级锁定效应（这导致阈值电压的巨大漂移）迟滞了高 k 介质的采用[26]。高 k 介质将很有可能首先用于对泄漏消耗要求非常严格的低功耗应用中。

2.1.3 应变工程

从 Stohr 和 Klemn[30] 以及 Wang 和 Alexander[31] 第一次制备得到均匀的 SiGe 合金到现在已有近 60 年了。Johnson 和 Christian 的开创性工作以及 Braunstein 等人关于单晶以及多晶 SiGe 合金的一系列文章为今天将 SiGe 引入先进的 CMOS 器件奠定了基础[33-36]。这些工作包括测量了锗在硅中摩尔百分比发生变化时，晶格常数和禁带宽度的变化。他们的结论显示，晶格常数从硅的 5.43 到锗的 5.66 之间呈现线性变化（基于后来结果的二次拟合）。在锗单晶与硅单晶之间约

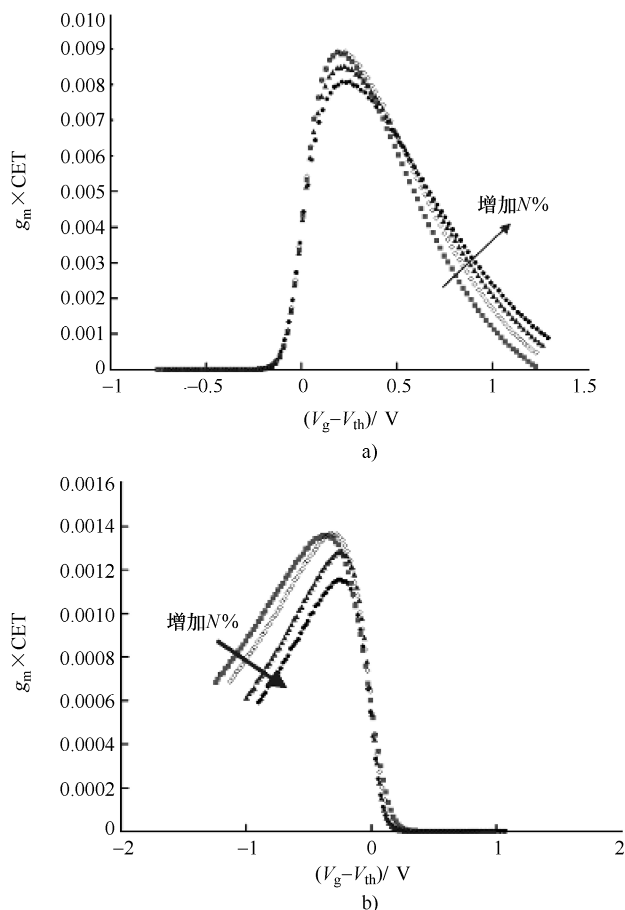


图 2-5　等离子氮化长沟道 a）NMOS 和 b）PMOS 器件的归一化跨导
（随着氮分的增加 NMOS 的高场跨导得到改善，而 PMOS 的高场跨导衰退）

4.2% 的晶格失配导致了 SiGe 合金电子能带结构的重大变化。不像晶格常数在整个组分范围内保持近似线性变化，Ge_xSi_{1-x}合金的能带开始以较小的斜率线性下降，而当硅中的锗达到 85% 时将以很陡峭的斜率下降。当硅在锗中的摩尔百分比下降到 15% 时，合金的导带由一个类似于硅的导带结构（E_g = 1.14eV）变到类似于锗的导带结构（E_g = 0.67eV），导致能带发生跳变。在合金组分的变化范围内，价带结构基本保持不变，其最大值在中心 k（000）处。合金的导带开始时类似于硅，其最小值沿着［100］晶向在 0.8 × 处。但是在硅中锗的组分达到 85% 时，能带最小值由类似于硅跳变到类似于锗。

在硅上 Ge_xSi_{1-x}的赝配淀积要求在生长方向上（［100］Si，见图 2-6）对晶格常数进行重大调整。在与生长方向平行的方向上，在硅中锗组分的整个变化范

围内，都要求晶格常数必须与硅的晶格常数保持一致。硅和锗晶格的金刚石结构变为在与生长方向平行的方向上存在明显压应力的四方结构，因此应变的程度与硅中锗的克分子百分比相关。由硅上 Ge_xSi_{1-x}的比例淀积带来的应变通过分割能带而改变了导带和价带的结构。例如，应变合金的价带被分割为重空穴和轻空穴两个能带。因为禁带宽度是价带顶与导带底的能量差，在硅中存在一定锗组分时，应变 Ge_xSi_{1-x}的禁带宽度将明显小于非应变合金的禁带宽度[38-41]。

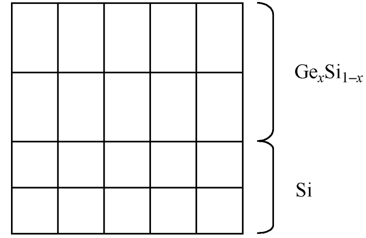


图 2-6 硅上 Ge_xSi_{1-x}的比例淀积导致 SiGe 合金晶格存在压应变（$Ge_{0.2}Si_{0.8}$的晶格常数比硅约大 1%，并且在硅上赝配生长的临界厚度可达几十纳米）

SiGe 合金禁带宽度变窄导致了异质结构 Si/Ge-Si 中硅和合金之间的能带偏移。在Ⅰ类能带排列中，对应硅上淀积 Ge-Si 合金，偏移发生在价带而导带基本对齐。而在Ⅱ类能带排列中，SiGe 上赝配生长硅，导带和价带都发生偏移。SiGe 合金的另一个重要特征是与硅相比具有较高的空穴迁移率，这是因为锗的空穴迁移率较高。另外，由于迁移率是散射和有效质量的函数，因此在应变 SiGe 合金中迁移率比非应变合金有进一步明显提高。较低的有效质量和较小的散射是由于能带图中简并度的提升[42-47]。

硅上 SiGe 的赝配生长和用于基础研究的异质结构都采用分子束外延。然而，在实际的生长中，SiGe 的淀积一般采用化学汽相淀积（CVD）的方法。典型的外延 CVD 系统的设计要考虑气氛和减压工艺。原子级的清洁表面是选择性淀积的关键，硅的表面一般都要采用稀释的氢氟酸进行预处理。在淀积之前，晶圆首先在炉内高温、通氢气的气氛下烘焙，去掉表面自然形成的氧化层。而淀积过程本身在较低的温度下进行，采用的温度取决于淀积采用材料的化学组成。一般用硅烷或者二氯硅烷作为硅源，用锗烷作为锗源，并且为了提高氧化物和氮化物的选择性，采用 HCl 气体与硅烷和锗烷混合在一起[48]。温度小于 800℃时在硅上可以获得一个近似理想的选择性 SiGe 淀积。独立区域温度控制的反应室设计与最佳灯加热方式的结合，再加上精确的温度控制和气流控制就可以实现均匀的淀积，且掺杂可高达 10^{21}。气体混合方式、温度和反应室设计的优化，以及掺杂气流与锗气流的分开，可以在整个晶圆上获得锗与掺杂浓度分布标准偏差（1σ）小于 1% 的均匀性。

2.1.4 快速热处理技术

快速热处理（Rapid Thermal Processing，RTP）是一项有着十余年历史的非常重要的半导体制造技术，它的发展历程可以回溯到 20 世纪 60 年代[49]。RTP

能取代传统的分批装片炉热处理有两个主要原因：优良的气氛控制和减小的热消耗。气氛控制是 RTP 得到广泛应用的最初动力。最初 RTP 工艺的应用是形成硅化锑（$TiSi_2$），正好使得 RTP 特长得到充分发挥。工艺中要求氮气气氛必须严格控制使环境氧气的水平小于 10ppm[㊀]。这种低氧含量在低流片量的单片 RTP 反应腔中很易实现，而在大流片量的批式炉管中则很难可靠地维持。早期 RTP 系统的缺点是温度控制和重复性不够好，但是生长 $TiSi_2$ 对温度控制和可重复性不很敏感，而 RTP 较短的工艺时间则具有明显优点，可以减小热退化和横向过度生长或桥接。

RTP 取代分批装片炉的第二个动力是可以减小晶圆在高温中的暴露时间。在高温氧化或退火时，加热炉中温度缓慢的上升和下降（典型的温度变化速率为 10℃/min）使晶圆长时间的处于高温下，这些过长的时间将导致不必要的掺杂和缺陷扩散。如图 2-7 所示，RTP 使温度上升和下降加速，典型的可以达到 75℃/s，这样就减小了晶圆过多的热暴露和不希望的扩散。

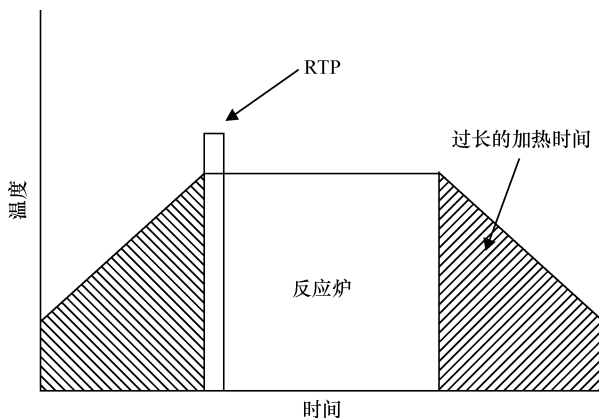


图 2-7　在 RTP 和批式炉之间的“热暴露”对比

RTP 技术　快速热处理要求设备具有特定的技术能力，包括精确的气氛控制、高达 250℃/min 的升温速度、峰值退火温度均匀性（3σ 为 3℃）、300 ~ 1200℃的温度控制范围和不受晶圆发射率影响的温度测量。在多数的 RTP 工序中都要求精确的气氛控制，包括金属的硅化处理、注入退火和热氧化。早期 RTP 设备的气氛控制依赖于大范围的净化，因为炉腔对生产环境是“打开的”，以装卸每个晶圆。这限制了生产量和最终能达到的气氛控制。一种改进的方式是将 RTP 腔装在集簇设备上，为实现纯净的氮气气氛，采用真空装载锁完成快速的气体交换。集簇设备也可以用于单腔或者多腔工序。气流模型的改进使非装载锁系

㊀ ppm，即 parts permillion，百万分之一。

统也可以在高产量的情况下达到低于 1ppm 的气体纯度。在腔门打开时，这些系统在晶圆输运过程中一般采用高速气体纯化以最大程度地避免氧气的侵入。

在注入退火时，为了达到要求的升温速度，RTP 要求由晶圆和直接支撑晶圆的结构所组成的整个被加热物体的“热质量”尽量得小。大多数 RTP 设备使晶圆与腔体之间热隔离，使用灯辐射方式满足快速升温时的热需求。另外一种方式是将晶圆快速地放入到一个其热质量比晶圆大得多的已加热环境中，此时晶圆引起的环境温度变化将很小，然后晶圆逐渐达到环境温度。

为了在 RTP 要求的较大温度范围内均匀性达到 3σ（在一个晶圆内和多个晶圆之间所有测量点的标准偏差的 3 倍）为 3℃的要求，需要对均匀性实现有效的控制。由于晶圆中心与边缘处表面积与体积之比不相同，在 RTP 加热过程中的不同部分要求不同的能量分布。如图 2-8 所示，在升温过程中，晶圆边缘的温度一般高于中心的温度，然而在温度稳定后和降温时，边缘的温度又会低于中心的温度。在采用灯加热方式的 RTP 设备中，解决这一问题的方法是在腔体的中心和边缘的不同区域采用不同的灯组合，并且在加热过程中调整不同灯的功率，保证不同时刻温度的均匀性。为实现积极实时的温度均匀性控制，最先进的系统采用高速、多点测量和多区域控制。

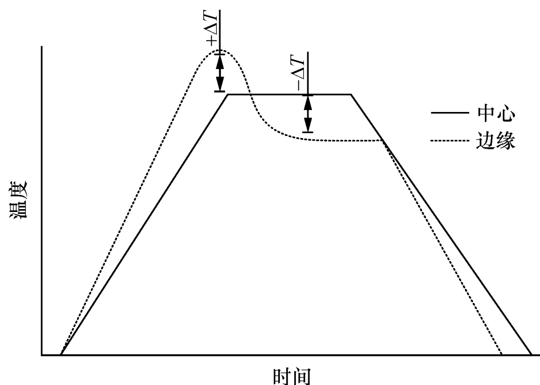


图 2-8　未能实现有效均匀性控制的中心和边缘温度分布

对均匀性控制最具挑战性的工艺是用于现代器件注入离子激活的高温尖峰退火。尖峰退火并不要求保持在峰值温度，而是到达预期的温度后马上降温，实现一个三角形或者尖刺形的温度-时间分布。注入离子的激活是与温度非常敏感的工艺，为了达到晶体管源漏区浅结的要求，需要实现 3σ 为 3℃的温度控制。图 2-9 所示为在 300mm 晶圆上距边缘为 3mm 范围内实现了上述温度均匀性的情况。

RTP 设备测温系统采用光学测温，它测量晶圆本身的灰体辐射以确定温度。在老式 RTP 系统中，用热电偶直接接触测量温度，但是这种方式由于热接触方式以及局部的热扰动会引起温度误差和不均匀性已被淘汰。准确的光学测量要求知道测量波长处晶圆发射率的数值（晶圆发射的能量与同样温度下理想黑体发射的能量之比）。测温装置设计采用两种主要的方式来应对这些挑战。一种方式是在晶圆下方创造一个高反射率的环境以实现理想黑体腔的特性，因为这能增强

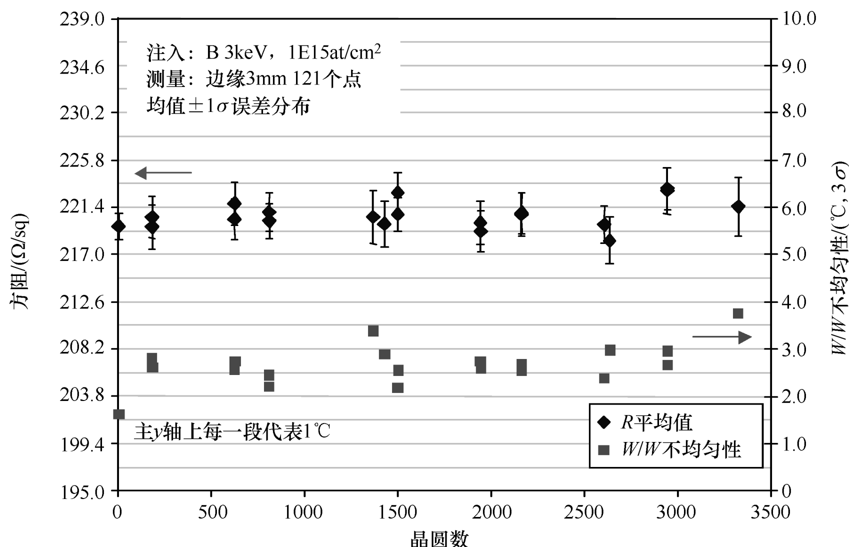


图 2-9　RTP 尖峰退火在 3σ 为 3.2℃控制时的方阻均匀性/可重复性

晶圆的有效发射率并减小可能出现的测量错误。另外，用一个分离的探针实时测量发射率并且对任何存在的误差进行校正[51]。第二种方式是实时测量灯功率周期变化的反射系数来获得假设无晶圆发射时的发射率（发射率 = 1 − 反射系数）。然后，用测量的发射率校正光学测温仪的读数[52]。

这些技术，包括气氛控制、低的热质量、均匀性控制和温度测量，共同构成了 RTP 设备的核心，并使 RTP 在很宽的温度范围内得到广泛的应用。对未来的 65nm 到 45nm 的器件，RTP 退火的热消耗即使是已经受到了限制，但对某些工艺步骤来说还是太高。一种方案是通过仅在晶圆的表面加热使暴露时间减小到毫秒级。这被称作热通量退火，可以用闪光灯或激光实现[53,54]。

RTP 应用　随着 RTP 工艺设备的改进和成熟，现在具备了用于完成 CMOS 和 DRAM 制造的所有的热退火和氧化工艺。主要的退火应用包括难熔金属硅化物的形成（例如 $TiSi_2$、$CoSi_2$、NiSi）、离子注入退火、BPSG 的稳定性和回流，以及接触金属退火（如 TiN）。主要的氧化应用包括栅氧氧化和氮化、隔离槽内壁隔离层氧化和牺牲氧化层氧化。

从 20 世纪 80 年代开始，自对准硅化钛技术被引入到 CMOS 工艺中，减小了源漏和栅区方阻，在晶体管与金属互连之间实现较好的欧姆接触。这道工艺是自对准的，因为在覆盖溅射金属后，金属仅在金属源漏和栅区位置直接与硅接触的地方形成硅化物，并不在侧墙隔离层和隔离结构上形成硅化物。近几年，为了进一步缩小器件尺寸，使用的金属已由钛变为钴再到镍。如图 2-10 所示，自对准金属硅化物一般经过两个 RTP 步骤形成。在清洗和溅射或物理汽相淀积（Physi-

cal Vapor Deposition，PVD）完成覆盖金属膜淀积之后，第一次 RTP 退火（通常称为 RTP1）形成金属硅化物。RTP1 的温度必须高到能使金属-硅发生反应并且生成物要达到足够的厚度和希望的金属硅化物相，但是温度又要足够低以免金属硅化物横向生长到栅的侧墙隔离层上，否则将导致栅与源漏短路。气氛控制对 RTP1 来说是关键，因为如果气体氛围中即使氧的比例只有百万分之几，金属薄膜也会很快被氧化耗尽。

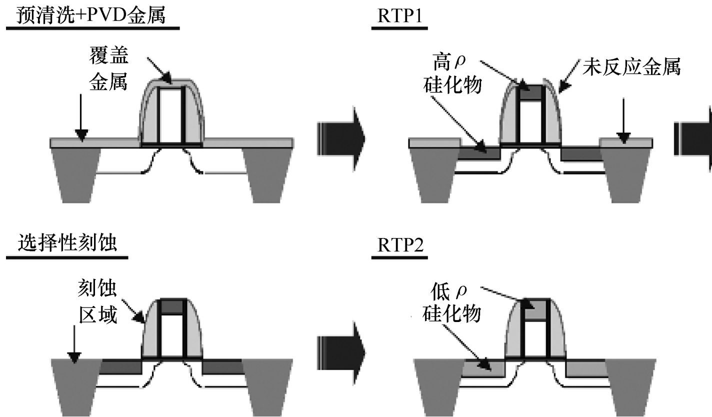


图 2-10 自对准工艺顺序

在选择性刻蚀移除侧墙隔离层和隔离区域上未反应的金属后，第二次退火（RTP2）通过改变金属硅化物相或组分来减小方阻。RTP2 一般要求温度高于 RTP1 几百度，这个温度太高以致于在二次退火这一步内不会生成金属硅化物。RTP2 的最高温度一般受到两个因素制约。一是温度太高会出现形态衰退，称作凝聚。另外如果温度太高可能使硅化物的相或结构发生变化，导致电阻退化。

钛的硅化物是首次在逻辑应用中被广泛使用的材料。$TiSi_2$的 C49 相一般在 RTP1 阶段，650℃、30s 的条件下形成。RTP2 将 C49 相硅化物变为 C54，可以实现大约（13～16）μΩ·cm 的低电阻率，RTP2 的工艺条件一般是 850℃，20s 或者更低。随着器件尺寸的缩小，使得钛硅化物的应用范围很难扩展到 0.25μm 以下。当线宽和金属硅化物的厚度进一步减小时，使 C49 到 C54 相变成核过程中不出现硅化物的凝聚将变得非常困难[55]。这些限制要求高温时间必须更短，而在高温时工艺窗口最宽。

为了改善缩小变窄的线宽，作为 $TiSi_2$的替代者，选用了钴硅化物。在钴自对准硅化物工艺中，RTP1 的工艺条件一般是 500～550℃、30～60s 形成 CoSi。RTP2 的工艺条件一般是 750～800℃、30s 使 CoSi 变为 $CoSi_2$，此时电阻率将降到 14～18μΩ·cm。在覆盖金属淀积期间，钴一般被钛或 TiN 覆盖以防止其氧化。

覆盖物将在 RTP1 与 RTP2 之间被选择性刻蚀掉。

在线宽缩小到 50nm 之下，钴硅化物也开始受到器件尺寸缩小的制约，并且驱使业界用镍硅化物取代之。一直到 30nm 线宽时，NiSi 还可以保持较低的方阻[56]。NiSi 可以用一步 RTP，而不是如图 2-10 所示的两步 RTP 工艺实现，RTP 温度一般在 400～500℃。然而，栅长小于 100nm 的方阻的减小表明，这会造成窄线条产生过量的硅化。采用两步 RTP 控制 NiSi 的形成可以解决过量的硅化和由多晶耗尽及结泄漏带来的问题。第一步，RTP1，形成富镍相的硅化物 Ni_2Si，而 RTP2 完成了实现低电阻率相 NiSi 的反应。两步形成 NiSi 要求 RTP 设备具有先进的能力。首先，RTP1 要求较低的工艺温度，250～350℃。这低于一般光学测量仪的测量范围。第二步，RTP1 要求精确地控制温度均匀性。硅化物的厚度是由反应温度而不是由淀积的镍的厚度决定的，并且硅化物生成反应对温度极其敏感[57]。最后，像之前的钛和钴一样，NiSi 要求精确地控制环境气氛，氧的含量要低于 1ppm。

CMOS 工艺主要依赖于离子注入和 RTP 退火以确定对器件性能有决定作用的杂质分布。热退火的作用是修复离子注入带来的损伤并使杂质原子处于替位状态，起到提供载流子的作用。由于 RTP 精确的热消耗控制，使得晶体管尺寸在横向和纵向的缩小成为可能，因此已取代了用于离子注入退火的批式加热炉工艺。如图 2-7 所示，由于减少了过多的热暴露时间，使 RTP 在实现与批式加热炉退火相同的方阻时，能实现更浅的结。

RTP 热消耗已经得到连续的减小，从原先 950～1050℃之间 30～60s 保温时间的退火变为 1050～1100℃的 0s 尖峰退火。将升温速度从 50℃/s 提高到大于 250℃/s，可以使 RTP 热消耗得到进一步减小。为了满足国际半导体技术路线图（Inter national Technology Roadmap for Semiconductors，ITRS）对源漏和其延伸的轻掺杂结的要求，热消耗还需要得到进一步减小。ITRS 规定了这些结的结深和方阻的目标，如图 2-11 所示。图中同时显示了 RTP 尖峰退火的低能注入耦合的典型性能。

虽然 90nm 的要求可以达到，但 65nm 要求结深和方阻进一步减小。几个技术的应用将使这种缩小成为可能，这些技术包括热熔退火、注入区的固态外延再生长，以及能在 RTP 退火时抑制扩散的离子同步注入。

几个因素能将 RTP 拓展应用到 CMOS 工艺中所有的热退火和氧化。首先，整个工艺流程中热消耗控制是关键器件尺寸能实现缩小的根本。第二，所有的氧化层厚度和退火步骤都已经缩小到单片加工工艺的范围。第三，新颖的工艺，例如现场水汽生成，已经使改善器件性能成为可能[59,60]。最后，单片工艺减小了周期时间，减少了工作量，使改善生产效率和生产时间有了可能[61]。

总之，RTP 在半导体中是一项关键的和促成性的技术。低温工序和毫秒级热

暴露的扩展能力将使器件的继续缩小成为可能。在未来的几年时间里，单片工艺的经济性和高效率将持续的促使热工艺从批式向 RTP 技术转换。

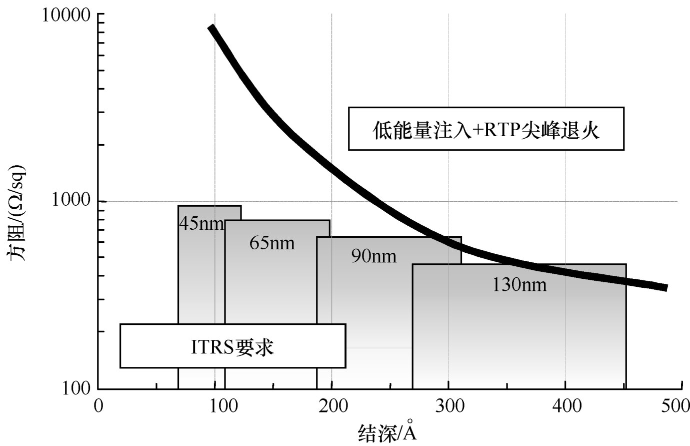


图 2-11　显示有 RTP 能力和 ITRS 要求的方阻与结深关系曲线

2.2　在 CMOS 尺寸缩小中与前道工序相关的器件问题

CMOS 工艺持续缩小到 sub-100nm 范围后，引起了大量在过去不太重要的器件物理和工艺方面的问题。因此，对 CMOS 技术，必须引入一些新的材料和新的器件结构，才能跟上国际半导体技术路线图预测的尺寸缩小步伐。本节我们将研究在逐渐缩小的 CMOS 器件中出现的与前道工艺相关的重要物理效应，并从器件和设计的观点出发来讨论一些新的工艺技术的影响。

2.2.1　CMOS 缩小的挑战

在过去的 30 年，传统的体硅 MOSFET 结构得到了成功的缩小。然而，在纳米时代，持续的缩小面临着基本的物理阻碍和经济性制约。下面简要分析影响 MOSFET 缩小的主要问题和因素。

短沟道效应　在理想的 MOSFET 中，栅压对沟道导通状态有着完全的控制。然而，当栅长变得更小时，漏区将对沟道电势影响更大。栅长非常短的器件，阈值电压 V_{th}将明显减小，导致关态泄漏激增，在漂移区，阈值电压的变化将变得非常大。为了抑制短沟道效应，必须减小漏与沟道之间的耦合，而增加栅与沟道之间的耦合。漏与沟道之间耦合的减小要求沟道和漏区有较小的耗尽区宽度，这可以由衬底的掺杂分布工程来完成。沟道逆掺杂分布（即掺杂浓度在表面最低，向衬底内部不断增加）和 halo 掺杂分布是减小耗尽区宽度的有效方法。源和漏

结也要求相应减小。由于 halo 掺杂有着很高的掺杂浓度，将在漏 – 衬结附近产生高场，导致能带间的隧穿引起较高的结泄漏。为了增加沟道到栅的耦合，必须采用较薄的栅氧电学厚度。

等效栅氧厚度缩小　较薄的栅氧，或者较大的单位面积栅电容，是改善关态和开态器件特性所必需的。现有最新技术使用了具有更好抗杂质穿通作用的氮化工艺，栅氧厚度已经减小到了 20Å 以下。对于这样薄的 SiO_2，即使电源电压随之减小，直接栅隧穿泄漏还是不能忽略。这引起了很严重的关态功耗问题。为了达到指定的等效氧化层厚度（EOT），如果采用高 k 介质，可以使用比 SiO_2 更厚的物理厚度，这是解决这个问题的潜在方案。当前，高 k 介质的一些问题依然在研究，例如迁移率衰退和平带电压漂移。了解了这些工艺技术的挑战后，必须开发出一些电路和系统级的技术以克服栅泄漏电流问题。具体分析不同应用类型的差异（即确定高性能、低工作功耗和低待机功耗应用），也有助于工艺技术与设计需求之间实现更好的相互匹配。

与栅介质 EOT 缩小相关的是多晶硅栅耗尽效应。当器件处于开态时，对典型的多晶硅掺杂浓度，在多晶硅中会存在一个薄耗尽层（约 1nm）。这个耗尽层削弱了栅到沟道的容性耦合，或者说减小了栅的过驱动电压。当栅介质的 EOT 接近 1nm 时，栅耗尽层的附加影响就变得很重要了。虽然增加栅的有效掺杂浓度很有帮助，完全解决这个问题的一种方式是使用金属栅电极。

沟道中的量子力学效应也给栅氧化层电容引入了一个附加的串联电容。当垂直电场足够高时，在当前和未来 CMOS 器件中就会出现下述情况，即沟道中载流子的垂直运动（与衬底表面垂直）将被限定在一个势阱中。与这个运动相关的能态将从经典情况下的连续态变为离散态，由此造成的一个影响是载流子分布的峰值位于离开衬底氧化层界面很小距离（约 1nm）的地方，这意味着给栅介质 EOT 又额外增加了几 Å 的厚度。与多晶硅栅耗尽效应一样，这也似乎是一个无关要紧的问题，直到 sub-100nm 技术节点，这个效应才需要给以关注。量子效应与几个因素有关，如垂直场、衬底掺杂和在超薄体 MOSFET 中硅体的厚度。这是一个基本的物理限制，不能简单地通过工艺改进解决。

沟道载流子迁移率　沟道载流子迁移率直接影响着电路和器件的性能。随着器件的缩小，有几个因素对迁移率产生不良影响。为了控制短沟道效应，通常采用的高沟道掺杂浓度，将导致更多的杂质散射，使迁移率降低。由于有些因素不可缩放，如热电压和硅的禁带宽度（禁带宽度影响阈值电压的设定，但是并没有被认为是影响缩放的基本障碍），电源电压的缩小比例跟不上器件尺寸的缩放比例。因此，沟道中载流子所处的平均垂直电场逐渐增加。根据迁移率模型，较高的垂直电场使迁移率衰退。由于 STI 工艺在沟道中产生了不希望的应力，因此也引起 n – MOSFET 迁移率的衰退。另外，大多数待选的正在评价中的高 k 介质

的界面都比热 SiO_2 栅介质的界面更差。较差的界面也使沟道载流子迁移率衰退并对使用高 k 介质的 MOSFET 器件的性能退化有直接的影响。

改善载流子迁移率有几个可能的方案。如前面所介绍的，ITRS 预测了沟道应变工程技术的引入。在一些新颖的器件结构中，如绝缘衬底上的超薄体硅（即 SOI）或者 FinFET，沟道掺杂极轻，这也有助于改善载流子迁移率。

工艺波动 工艺波动对设计而言是一个需要考虑的非常重要的问题。随着器件的缩小，关键器件参数如阈值电压 V_{th} 的波动将增加。引起波动的原因很多，主要来源包括掺杂源的波动、氧化层厚度的起伏、特征尺寸（CD）的起伏和线条边缘的不平直。对较小的器件尺寸，沟道中掺杂总量降低，因此统计波动上升了 $1/N^{1/2}$，这里 N 是杂质总量。虽然最终的 V_{th} 变化依赖于杂质分布，但是从阈值电压的角度而言，将影响器件的进一步缩小。一些要求沟道掺杂浓度非常低的新颖的器件结构，如双栅 MOSFET，可以用来解决这个问题。但是这些器件可能对其他的波动源很敏感（如硅体厚度）。随着器件尺寸减小，CD 和薄膜厚度变化的容差也随之变小。除了工艺改进以实现更好的成品率和性能之外，需要采用更多的鲁棒性设计方法。

新颖的器件结构 虽然体硅 MOSFET 的栅长被证实可以降到 15nm[64]，但是 CMOS 的最终缩小将可能需要采用新颖的器件结构。最近几年，FinFET 及其几种多栅结构在业界和科研中已被证实可以成功地实现尺寸缩小。超薄体 SOI MOSFET 也是极有前途的选项。缩小器件尺寸的实用解决方法必须是能以性能改进或者减小每项功能开销的形式提供较好的收益。这将要求以一种经济的方式整合各种新颖的工艺和器件革新。

2.2.2 量子效应模型

以前，栅氧厚度是栅-沟道电容中支配性的因素。然而，随着栅氧厚度缩减到 20Å 以下，其他两个因素变得重要。一个因素与沟道中反型层载流子的量子限制效应相关。另一个是在反向栅偏置时栅耗尽层电容。

直接影响关态电流的热电压不能缩小。因此，阈值电压必须足够大以使关态泄漏电流可以接受。再结合考虑到驱动电流的要求，就限制了电源电压的缩小。结果是沟道和氧化层中的垂直电场随着器件的缩小而增加，衬底氧化层界面的强垂直电场形成了一个势阱，这将限制反型层载流子在垂直方向的运动。在垂直运动方面，载流子将变成具有离散本征能级的二维电子气。另外，反型层载流子的峰值分布取决于不同能带中所有载流子的波函数，其峰值离衬底氧化层界面 1 ~ 2nm，从而在栅电极下产生额外的电容（见图 2-12）。当栅介质 EOT 缩小到 20Å 以下时，此电容不再可以忽略。

反型层电荷分布的严格分析处理要求对氧化层和衬底区中泊松方程和薛定锷

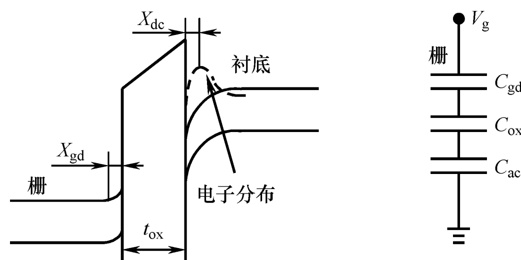


图 2-12 额外的电容

（左图为反型偏置的 n⁺ 多晶硅栅 n 沟道 MOSFET 的能带图沟道电子位于衬底界面附近势阱的离散能带上。电荷中心与衬底界面距离为 X_{dc}（dc 电荷中心）。多晶硅栅的耗尽宽度是 X_{gd}（在 2.2.3 节讨论）。右图为栅电容层叠的交流等效电路。C_{gd}，C_{ox} 和 C_{ac} 分别是栅耗尽电容、栅氧电容和反型电荷 ac 电荷中心带来的电容）

方程自洽求解，这需要采用数值计算方法。反型层载流子分布取决于栅电压和器件参数，如沟道掺杂浓度和栅氧厚度。基于数值仿真结果和实验数据，提出的一个经验模型可以对反型电荷 ac 电荷中心有一个简单的精度较好的估算[65]。对于薄氧化层和典型沟道掺杂水平情况，在 ac 电荷中心 X_{ac}、栅电压 V_g、阈值电压 V_{th}、和栅氧 EOTt_{ox}之间存在如下普遍关系：

$$X_{ac}=6.20\times10^{-5}\left(\frac{V_g+V_{th}}{t_{ox}}\right)^{-0.4} \tag{2.2.1}$$

式中，V_g和 V_{th}的单位是 V（伏特），而 X_{ac}和 t_{ox}的单位是 cm（厘米）。图 2-13 所示为式（2.2.1）表示的关系。在 90nm 节点之下，X_{ac}的范围是 8 ~ 10Å，使反型层电容的等效厚度（CET）增加 3Å。对大约 10Å 的 EOT 的栅介质，这将明显地引起栅电容的减小。

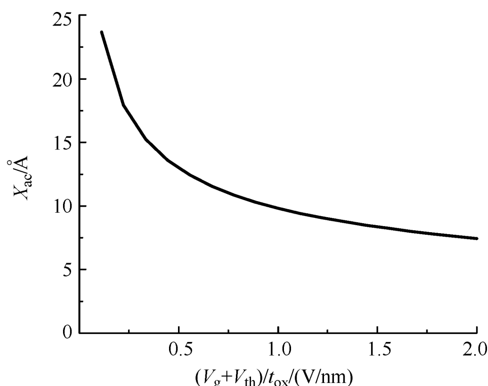


图 2-13 ac 电荷中心 X_{ac}与（V_g+V_{th}）/t_{ox}之间的普遍关系

上面模型是基于传统的体硅 MOSFET 结构。在一些新颖的器件结构中，如

双栅 MOSFET 或超薄体 SOI MOSFET，由于边界条件的不同，描述数值结果的公式可能会不同。

2.2.3 多晶硅栅耗尽效应

当栅氧中存在电场时，静电边界条件使多晶硅靠近氧化层界面附近的能带弯曲。当栅氧足够薄时，硅栅和沟道之间的耗尽区所贡献的电容就需要特别注意了。图 2-12 显示了反型区栅电容各分量的等效电路。等效栅电容（即 CET）与这些电容分量的关系为

$$\frac{1}{C_g}=\frac{1}{C_{gd}}+\frac{1}{C_{ox}}+\frac{1}{C_{ac}}=\frac{X_{gd}}{\varepsilon_0\varepsilon_{Si}}+\frac{t_{ox}}{\varepsilon_0\varepsilon_{ox}}+\frac{X_{ac}}{\varepsilon_0\varepsilon_{Si}} \tag{2.2.2}$$

式中，变量的含义与图 2-12 中一致。$\varepsilon_0=8.85\times10^{-12}$ F/m，是真空介电常数；$\varepsilon_{Si}=11.7$，是硅的相对介电常数。在式（2.2.2）中，栅介质的物理厚度为 t_{ox}，相对介电常数为 ε_{ox}。栅介质的 EOT 由下式给出：

$$\text{EOT}=t_{ox}\frac{\varepsilon_{SiO_2}}{\varepsilon_{ox}} \tag{2.2.3}$$

式中，$\varepsilon_{SiO_2}=3.9$，是 SiO_2 的相对介电常数。因此，从电容角度考虑，厚度为 t_{ox} 的栅介质与厚度为 EOT 值的 SiO_2 膜等效。通常用 CET 描述总的有效栅电容

$$\text{CET}=\frac{\varepsilon_0\varepsilon_{SiO_2}}{C_G}\approx\text{EOT}+\frac{X_{gd}+X_{ac}}{3} \tag{2.2.4}$$

X_{gd} 和 X_{ac} 是与偏置有关的，因此 CET 也依赖于栅压 V_G。在典型的器件工作条件下，栅耗尽和量子效应对 CET 分别贡献有几 Å 的厚度，所以使用超薄栅介质时，CET 将明显大于 EOT（按百分比计算），具体大小由栅介质决定。较薄的 CET 和较高的栅压将转化为较高的反型电荷密度和较好的器件特性，因此，随着每一代工艺技术中采用的电源电压的减小，栅层叠结构的缩小不仅要求栅介质（EOT）变薄而且要求栅耗尽效应缩小（X_{gd}）。

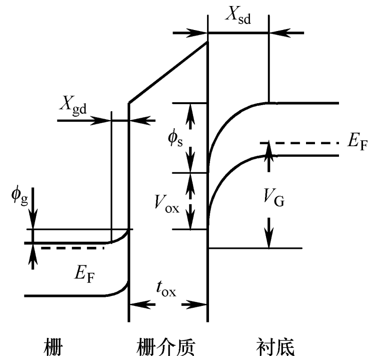


图 2-14 n – MOSFET 栅正偏置时的能带图
（靠近氧化层界面的多晶硅栅中存在着有限厚度的耗尽层）

采用静电学可以简单估算栅耗尽效应和 CET。考虑一个偏置到反型区的 n – MOSFET 的 n + 型多晶硅（见图 2-14），平带电压（V_{FB}）和衬底电压降（ϕ_s）、栅电压降（ϕ_g）、氧化层电压降（V_{ox}）之和等于栅压 V_g。

$$V_g = V_{FB} + V_{ox} + \phi_s + \phi_g \tag{2.2.5}$$

栅的掺杂浓度一般高于沟道掺杂浓度。因此在亚阈区，氧化层界面电位移的连续性意味着栅极的能带弯曲小于衬底的能带弯曲。在强反型条件下，栅氧上界面的边界条件是

$$\varepsilon_{ox}\frac{V_{ox}}{t_{ox}} = \varepsilon_{Si}E_{gate} \tag{2.2.6}$$

使用耗尽近似和假定栅的掺杂浓度（N_{gate}）为均匀掺杂时，栅极电场 E_z 可以使用高斯定理求解

$$\frac{dE_z}{dz} = \frac{qN_{gate}}{\varepsilon_0\varepsilon_{Si}} \tag{2.2.7}$$

因此栅的电压降为

$$V_{ox} = \gamma_G\sqrt{\phi_g} \tag{2.2.8}$$

式中，$\gamma_G \equiv \sqrt{2\varepsilon_0\varepsilon_{Si}N_{gate}}/C_{ox}$，其中 $C_{ox} = \varepsilon_{ox}\varepsilon_0/t_{ox}$。

在强反型开始前，大约在 $2\phi_b = 2(k_BT/q)\ln(N_{sub}/n_i)$ 时，衬底能带弯曲 ϕ_s 达到饱和。注意

$$V_{th} = V_{FB} + 2\phi_b + \gamma_S\sqrt{2\phi_b} \tag{2.2.9}$$

式中，γ_S 的定义与 γ_G 的定义相似，仅用 N_{sub} 代替 N_{gate}。因此式（2.2.5）变为

$$\phi_g + V_{ox} - (V_g + \gamma_S\sqrt{2\phi_b} - V_{th}) = 0 \tag{2.2.10}$$

当 $V_g > V_{th}$ 时，栅耗尽宽度 X_{gd} 可由式（2.2.6～式 2.2.10）求出

$$X_{gd} = \frac{\varepsilon_0\varepsilon_{Si}}{C_{ox}}\left[\sqrt{1 + \frac{2C_{ox}^2}{\varepsilon_0\varepsilon_{Si}qN_{gate}}(V_g - V_{th} + \gamma_S\sqrt{2\phi_b})} - 1\right]$$
$$\approx \frac{C_{ox}(V_g - V_{th} + \gamma_S\sqrt{2\phi_S})}{qN_{gate}} \tag{2.2.11}$$

对于给定的栅压、氧化层厚度、栅掺杂浓度和阈值电压，结合电荷质心模型，可以由式（2.2.4）和式（2.2.11）估计 CET。例如，假定 EOT = 12Å，$V_G = 1.1$V 和 $V_{th} = 0.3$V，由式（2.2.1）得到 ac 电荷中心 $X_{ac} = 9.2$Å。进一步假定 $N_{sub} = 3 \times 10^{17}\text{cm}^{-3}$ 和 $N_{gate} = 1 \times 10^{20}\text{cm}^{-3}$，可以估算得到栅耗尽层宽度 $X_{gd} =$ 15.7Å。因此，量子效应和栅耗尽效应一起对 CET 的贡献为 8.3Å，而栅介质的贡献为剩下的 12Å。显然，对于很薄的 EOT，再缩小栅介质对于减小 CET 作用已经不大了，因为多晶硅栅耗尽和量子效应的作用在 CET 中占的百分比已经很高。因为量子效应的贡献不能被减小，因此通过工艺改进或者使用新颖的栅电极材料（如金属栅电极）来减小多晶栅耗尽效应就变得非常关键。国际半导体技术路线图（ITRS）预测，在未来几年，量子效应和栅耗尽效应会给 EOT 附加 8Å，而在 2007 年，65nm 中将采用金属栅，它会给 EOT 附加 5Å[66]。

2.2.4　金属栅电极

多晶硅栅耗尽问题在几代工艺中都是制约器件性能改善的瓶颈。如前面讨论所指出的，多晶硅栅不能符合最新的 ITRS 要求，使用金属栅电极将是解决这一问题的方案。通常金属栅比金属硅化物多晶硅栅有更低的栅阻抗，这将有助于减小 *RC* 时延，这是采用金属栅的又一个好处。

使用金属栅电极对工艺和集成提出了许多挑战。从器件和设计的角度看，栅的功函数是主要的关注点。沟道掺杂适当时，使用 n+/p+多晶硅电极可以很容易地获得合适的 n-和 p-MOSFET 阈值电压。对于体硅 CMOS，使用金属栅很难满足同样大小栅功函数的要求。一般地，功函数高（类似 P+Si-）的金属是非活性的，因此很难刻蚀掉，而那些功函数低的金属又活性太高，与栅介质接触时会带来热稳定性问题。在体硅 CMOS 中使用功函数大小适中的金属时，将导致 n-和 p-MOSFET 的阈值电压大大高于希望值。但是，还是可以在两种类型的器件中使用同一种金属，然后分别调整两种类型器件上的金属功函数以优化这两类器件的阈值电压。当前，一种可接受的金属栅方案仍在研究中。除了获得合适的功函数，功函数分布的严格控制也是金属栅技术中非常重要的问题。

高 k 栅介质的使用影响着阈值电压的精确设定。许多高 k 介质具有高界面陷阱密度和固定电荷，这将引起显著的 V_{th}漂移和较大的 V_{th}变化。另外，当高 k 栅介质与金属栅一起使用时，将发生一项重要的物理效应。首先通过实验观测到，当使用不同栅介质时，p-MOSFET 的钼栅会有不同的栅功函数值（见图 2-15）。这可以用这些介质的界面偶极子具有不同的屏蔽效应来解释[68]。多晶硅栅很少受这种效应的影响，因为在禁带宽度中的能带密度可以忽略。理论模型预测，在高 k 介质上实现与 n+/p+硅类似的功函数，需要金属功函数有非常大的范围，这给候选的金属电极施加了一个更加严格的限制。

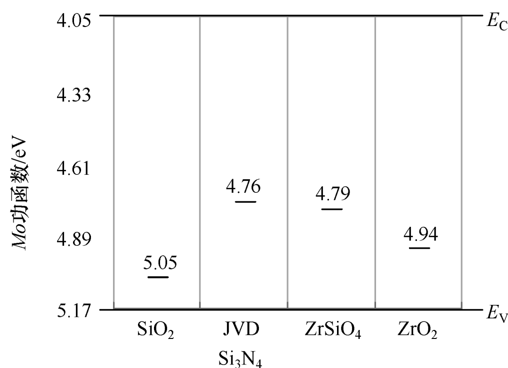


图 2-15　采用 p-MOSFET 测得的在不同栅介质上的钼栅的功函数视在值

（功函数随钼栅下方的栅介质不同而不同，并与真空值不同）（来源于参考文献［67］）

对于沟道掺杂浓度非常低的器件，比如 FinFET 或超薄体 SOI MOSFET，合适的 V_{th}要求栅功函数更靠近硅禁带中央。因此，候选金属的选择会稍稍容易。据仿真预测，对于 FinFET，偏离硅禁带中央 ±0.2eV 的栅功函数是适合于 p-和 n-MOSFET 器件的[69]。为了实现使用金属栅得到相对较小的功函数范围，有几项技术可以被采用，例如，掺杂的镍硅化物栅、在钼中注入氮，以及金属混合及合金。然而，一个完整的方案还有待研究。

2.2.5　栅直接隧穿泄漏电流

对超薄栅氧，即使在低栅压下，由于直接隧穿过程，会有相当的栅泄漏电流存在。随着器件的缩小，考虑功耗问题时，栅泄漏电流就变成了日益严重的问题。在 65nm 工艺中，EOT 接近 1nm，要求栅泄漏电流比 1nm SiO_2泄漏电流小几个数量级（随着应用对象的不同有所不同）。因此，作为 SiO_2栅氧的替代物，高 k 栅介质的研究正紧锣密鼓地进行。

在低于 20Å 的情况下，常规的硅热氧化物不能再作为电子和空穴的有效势垒，因而会引起了不可接受的高栅泄漏电流。不像 Fowler-Nordheim 隧穿机制那样，对于直接隧穿电流没有一个简单的解析方程。另外，沟道载流子的量子制约效应在计算沟道电流时不能被忽略。因此，隧穿电流的严格的物理模型包括自洽求解薛定锷方程和泊松方程以及计算不同量子态下载流子的分布[70]。另一方面，一个相对简单的解析模型，对于电路仿真和栅泄漏的快速估算非常有用。Lee 等人提出了一个半经验的 SiO_2直接隧穿模型[71]。一般情况下，需要采用下述三个直接隧穿分量组成总的栅泄漏电流模型：电子导带（ECB）隧穿、空穴价带（HVB）隧穿和电子价带（EVB）隧穿（见图 2-16）。随着偏置条件的不同，一些机制可以被忽略。例如，当氧化层电压 $|V_{ox}|$小于硅带隙电压 1.12V 时，不会

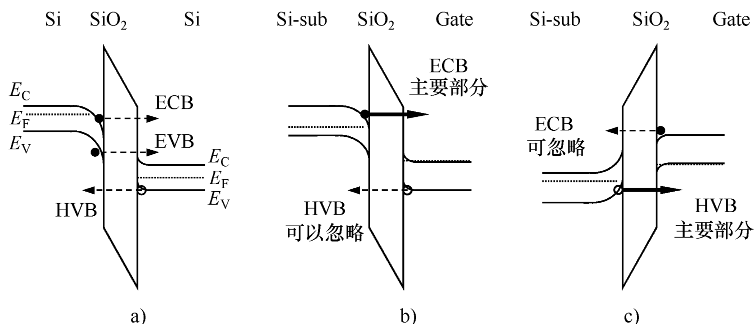


图 2-16　a）一般情况下，建立穿过薄栅氧的直接栅隧穿电流模型需要三个分量。在低栅压下，b）对 n-FET 主要贡献是 ECB 隧穿，c）对 p-FET 主要贡献是 HVB 隧穿（来源于参考文献［71］和［72］）

发生 EVB 隧穿。在低 V_G（<1.5V）情况下，对于 n-FET，反型偏置栅泄漏主要分量由 ECB 贡献，而对 p-FET 时，则是 HVB。

对每种隧穿机制，直接隧穿电流可以由下式表示：

$$J = \frac{q^3}{8\pi h\phi_b\varepsilon_{ox}} C(V_G, V_{ox}, t_{ox}, \phi_b)\exp\left\{-\frac{8\pi\sqrt{2m_{ox}\phi_b^3}}{3hq|E_{ox}|}\left[1-\left(1-\frac{V_{ox}q}{\phi_b}\right)^{3/2}\right]\right\} \tag{2.2.12}$$

式中，指数项代表在导带边界载流子隧穿概率的 WKB 近似，而 C 基本上是为了在低栅压范围能正确描述的 $J_g - V_g$ 变化关系的经验形状因子。

$$C(V_g, V_{ox}, t_{ox}, \phi_b) = \exp\left[\frac{20}{\phi_b}\left(\frac{|V_{ox}|-\phi_b}{\phi_{b0}}+1\right)^{\alpha}\left(1-\frac{|V_{ox}|}{\phi_b}\right)\right]\frac{V_g}{t_{ox}}N \tag{2.2.13}$$

式（2.2.13）中的 N 与隧穿载流子密度相关。对反型区和积累区，N 的一般表达形式为

$$N = \frac{\varepsilon_{ox}}{t_{ox}}\left\{S\ln\left[1+\exp\left(\frac{V_{ge}-V_{th}}{S}\right)\right]+v_t\ln\left[1+\exp\left(-\frac{V_g-V_{fb}}{v_t}\right)\right]\right\} \tag{2.2.14}$$

在式（2.2.14）中，S 是亚阈值摆幅；v_t 是热电压；V_{th} 是 MOSFET 的阈值电压。

该模型中的拟合参数是有效质量 m_{ox}、势垒高度 ϕ_b 和 ϕ_{b0} 以及 α。对不同的栅偏置和栅与衬底掺杂类型，选用合适的参数都会使模型与隧穿电流符合。对于上述隧穿的每个分量，模型参数是不同的。图 2-17 所示为该模型的精度。图中虚线是数值仿真的结果[70]、实线表示解析模型计算结果，而实心点符号是实际器件的测量数据。可以看到三种结果之间符合得很好。除了 SiO_2，上述模型也在单层非 SiO_2 介质应用中得到证明[72]。业已表明，对于采用喷射汽相淀积（Jet

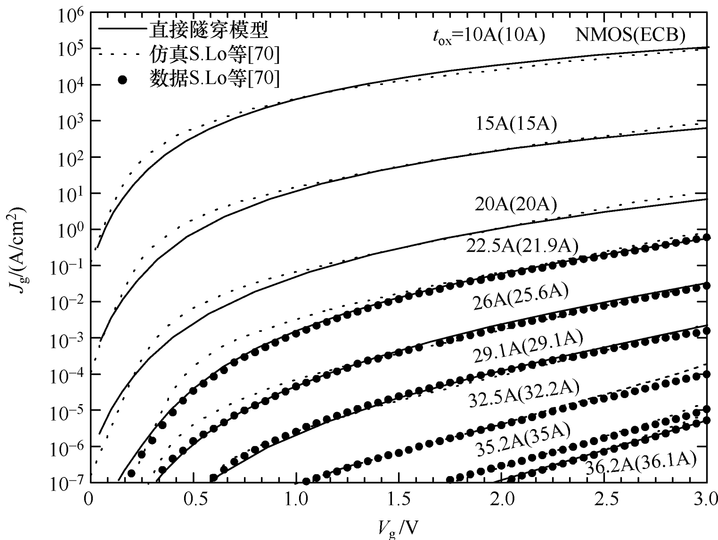


图 2-17　Lee 与 Hu[71] 的直接隧穿模型与实验数据以及仿真结果符合得很好

Vapor Deposited，JVD）技术生长的氮化硅，直接隧穿泄漏也可以与使用单独一组拟合参数的这个模型符合得很好。表 2-1 总结了在使用 SiO_2 和 JVD 氮化硅栅介质的 n-和 p-FET 中主要隧穿机制的参数。势垒高度可以通过其他方式获得，因此，实际上仅仅有效质量和 α 需要通过实验栅泄漏数据确定。为了更适合于有效的数值实现，这个模型的概念经过适当修改，被包括到 BSIM4 栅泄漏模型中[73]。

表 2-1　SiO_2 和 JVD 氮化硅栅介质的直接隧穿模型参数

	SiO_2		JVD 氮化硅	
	ECB	HVB	ECB	HVB
ϕ_b	3.10	4.50	2.10	1.90
ϕ_{b0}	3.10	4.50	2.10	1.90
m_{ox}	0.40	0.32	0.50	0.41
α	0.6	0.4	0.4	1.0

应该注意到，使用一些高 k 栅介质时，在高 k 栅介质和硅衬底之间存在一个薄界面层，这既可以是为了获得更好的薄膜质量而有意引入的，也可能是在高 k 层和衬底之间由不希望的化学反应生成的。因此栅绝缘层实际上包括了两个或者更多层不同的介质。由于上述模型是基于单层的概念，并不能被应用到多层介质结构。最近的一项工作将 BSIM4 中的单层隧穿模型拓展到了多层情况下[74]。结果表明，通过对多层中的每一层构建合适的用 BSIM 模型里隧穿参数表示的隧穿概率，BSIM 直接隧穿模型也与多层栅介质的实验数据符合得很好。目前，对于明确知道层结构的多层隧穿的实验栅泄漏数据非常有限。随着高 k 栅介质工艺技术的进一步改进，模型也需要进一步提炼。

2.2.6　寄生电容

栅与源/漏之间的电容是影响 CMOS 器件和电路性能的重要因素。栅和源-漏之间的耦合，一部分由于直接交叠产生，其不能随着普通工艺技术的缩小而成比例的缩小。因此，对于更短栅长的器件，寄生电容变得更加严重。隔离层介质材料影响着边缘电容。另外，用高 k 栅介质取代 SiO_2 栅氧化层也影响着寄生分量。虽然非常准确地计算电容要求二维数值计算，但是一个简单的物理模型可以被用来理解这些机制[75]。如图 2-18 中三个箭头所示，栅到漏（或源）电容包含了三个耦合途径：两个边缘分量和栅到漏的直接交叠。侧墙区两侧的边缘电容为

$$C_1 = \frac{\varepsilon_{SPCR} W}{\alpha} \ln\left(1 + \frac{t_G}{t_{ox}}\right) \tag{2.2.15}$$

式中，W 是器件的宽度；t_G 和 t_{ox} 分别是栅介质和栅的物理厚度；α 是如图 2-18 中所示的角度；ε_{SPCR} 是隔离层介质的介电常数。

直接交叠电容等于

$$C_2 = \frac{\varepsilon_{ox} W(d + \Delta)}{t_{ox}} \quad (2.2.16)$$

式中，d 是交叠长度；当 α 是 $\pi/2$ 时，Δ 是零。结侧面的耦合电容等于

$$C_3 = \frac{\varepsilon_{ox} W}{\beta} \ln\left(1 + \frac{x_j \sin\beta}{t_{ox}}\right) \quad (2.2.17)$$

式中，$\beta = \pi\varepsilon_{ox}/2\varepsilon_{Si}$。注意，当栅下存在自由载流子时，耦合途径将被屏蔽，因此在强反型时，这部分不存在。在上面几个表达式中，所有的介电常数值的单位是 F·cm。

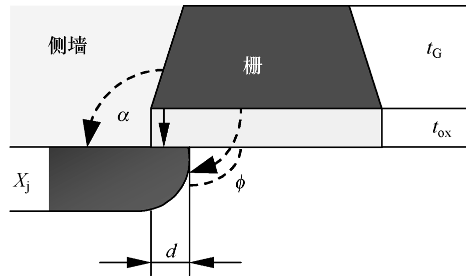


图 2-18　MOSFET 的栅与漏之间耦合路径简图（三个耦合路径均用箭号标出，器件宽度方向与纸面垂直）（来源于参考文献［75］）

C_1与器件宽度和侧墙介质的介电常数有着线性关系，而与栅和栅氧之间厚度之比的关系较弱。假定 $\alpha = \pi/2$，$\varepsilon_{SPCR} = 3.9\varepsilon_0(SiO_2)$，$t_{ox} = 12Å$，$X_g = 1000Å$，可以得到 $C_1 = 0.097fF/\mu m$。更厚的物理栅介质可以稍稍减小 C_1。如果使用高 k 介质，使得 $t_{ox} = 48Å$（EOT 仍然是 12Å），而其他一切都与上面例子中的一样，C_1 只减小到 0.068fF/μm。C_1减小的关键是确保侧墙介质有较低的介电常数。如果用氮化物侧墙取代 SiO_2，由于 $\varepsilon_{SiN} \approx 2\varepsilon_{SiO_2}$，边缘电容将增加到大约为原来的 2 倍。如果在栅介质刻蚀期间，高 k 介质不能从侧墙区的底部完全清除，C_1也会显著增加。

直接交叠电容 C_2随着栅沟道电容 C_{ox}缩小而线性缩小，其主要变量是交叠长度 d。希望 d 尽量小，因为这样可以减小交叠电容。然而，结果显示，为了避免短沟 MOSFET 中驱动电流的衰退，d 必须为 15～20nm[76]。如果源漏扩展（SDE）深度或 SDE 到栅的交叠太小，会增加沟道的串联阻抗，引起漏电流衰退。图 2-19显示了减小 Miller 电容（$C_M = C_1 + C_2$）和改善驱动电流 $I_{d\,sat}$之间的折衷。为了避免驱动电流衰退，要求交叠不能小于一个最小值意味着随着器件的缩小，交叠电容相对于栅电容 C_{ox}将变得更大。假定栅介质 EOT = 12Å，最小交叠为 15nm，会有 0.43fF/μm，而对于 60nm 的沟道长度，栅沟道电容是 1.73fF/μm。

寄生电容的另一个来源是源漏结电容。p-n 结存在耗尽区。当结偏置电压改变时，耗尽宽度随之变化，从而对电容负载产生影响。在 MOSFET 中，这些结的一侧一般都是重掺杂，因此耗尽宽度主要由轻掺杂一侧决定：

$$C_j = \frac{\varepsilon_{Si}\varepsilon_0}{X_D} \approx \sqrt{\frac{\varepsilon_{Si}\varepsilon_0 N_{sub}}{2(\phi_{bi} + V_R)}} \quad (2.2.18)$$

式中，X_D是耗尽层宽度；N_{sub}是衬底掺杂浓度；V_R是结上的反偏电压；ϕ_{bi}是 p-n 结内建电势。式（2.2.18）给出的是单位面积结电容。为了使总电容最小，保持结区下衬底掺杂浓度尽可能低就显得非常重要。另外，电路版图中结面积也应该尽量小。

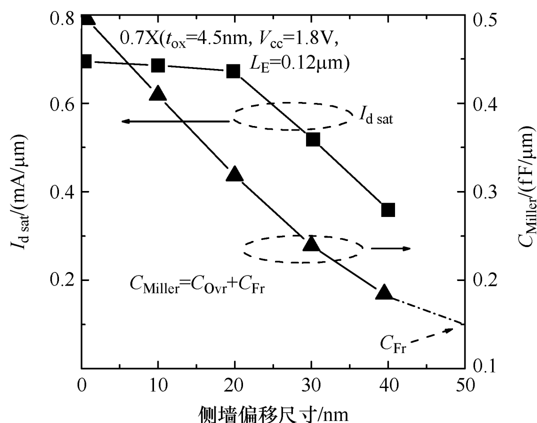


图 2-19　对给定的工艺，侧墙偏移尺寸对 Miller 电容（交叠电容加边缘电容）与饱和漏电流的影响[76]

（实验表明，用来控制 SDE 注入与栅边界之间距离的侧墙偏移尺寸为 0 ~ 40nm，使得 SDE 到栅的交叠可以在 40 ~ 0nm 范围调节。当交叠长度低于 15 ~ 20nm 的最小值时，漏电流就会发生退化）

高 k 栅介质会增加一个附加的耦合通道，而这在 SiO_2 中不重要。如图 2-20b 所示，当栅介质的介电常数很高时，为了得到给定的 EOT，它的物理厚度会远大于相应的 SiO_2 的厚度。漏和源的边缘电场通过厚栅介质的侧面影响着沟道。附加的耦合使短沟器件特性衰退。通过仿真对比了图 2-20 所示 3 种情况下 V_{th} 的漂移特性。对于相同的 EOT，图 2-20 中栅介质分别设定为：图 2-20a 是 SiO_2；图 2-20b 是高 k 介质；而图 2-20c 是在一个分界层上有高 k 介质[77]。假定器件的参数相同，1.5nm 的 EOT 可以采用 1.5nm 的 SiO_2，也可以采用 $k=200$ 的介质，或者是在 1.0nm 的 SiO_2 上采用 $k=200$ 且 EOT = 0.5nm 的介质来实现。图2-21所示为阈值电压与栅长的函数关系。由图可见，当使用 k 值非常高的单层介质时，

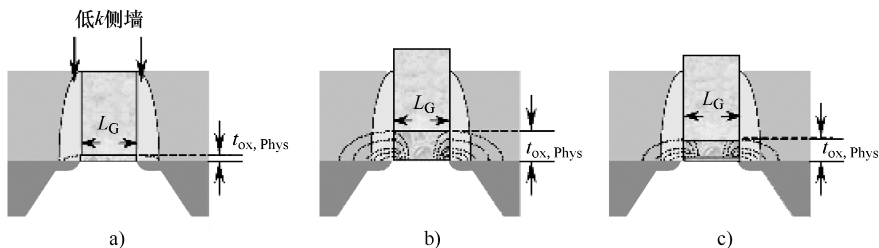


图 2-20　3 种情况下 V_{th} 的漂移特性

（对于给定的 EOT 值，单层高 k 介质（图 b）的物理厚度比 SiO_2（图 a）更大，一个更现实的栅介质结构是在一个薄的低 k（如 SiO_2）界面层上加一层高 k 介质（图 c）。对超薄 SiO_2 来说，源-漏与沟道之间的边缘电场是可以忽略的，而使用高 k 介质的情况，边缘电场增强）

V_{th}的漂移特性衰退很严重。而采用低 k 值的薄分界层可以有效避免这种不希望的作用。仿真结果也表明，即使 EOT 不同，如果物理 t_{ox} 与栅长之比相同，也会产生相同量的短沟性能退化。应该注意到，当前最有希望的高 k 介质也比仿真中假定的 k 值要低得多。因此，除非栅长变得小于 40 ~ 50nm 使 t_{ox}/L_G将变得足够大，否则这个效应不是很重要。总之，从短沟道效应的角度考虑，即使非常高的 k 值很有用，但是也是不希望的。

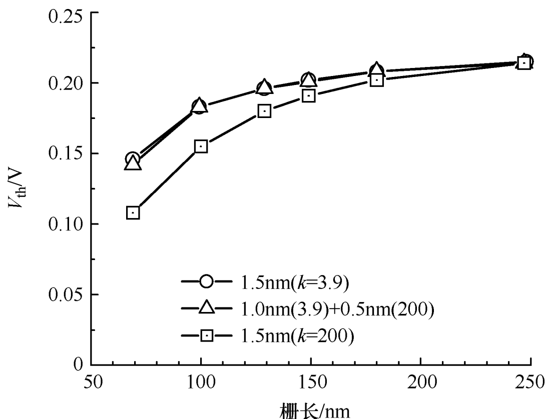


图 2-21 阈值电压与栅长的函数关系
（图 2-20 所示的 EOT 均为 1.5nm 的三个栅层叠结构的 V_{th}偏移特性。对于高 k 值单层栅介质，由于栅介质侧面增强了漏 - 源与沟道之间的耦合，V_{th}发生了明显的衰退。采用薄的低 k 界面层可以有效地防止退化）

从器件的角度看，使用高 k 介质的唯一好处就是减少直接隧穿栅电流。目前，高 k 介质存在大量的器件和工艺问题，例如迁移率衰退、平带电压漂移等。不管可制造的高 k 方案的可用性，功耗问题必须通过多个领域的前沿技术来解决，包括从器件和工艺到电路和系统设计。

2.2.7 需要关注的可靠性问题

器件的可靠性是 CMOS 工艺需要考虑的关键问题。更小的器件尺寸和新材料的使用，如铜/低 k 介质、高 k 栅介质和金属栅电极，都对 CMOS 器件可靠性有着重要的影响。集成规模不断增大对电路中的每个单元提出了更加严格的可靠性要求。测试低失效率复杂电路的开销和时间也是一个严重的挑战。下面简要分析与前道工艺相关的 CMOS 可靠性问题。

经时介质击穿 在过去的几十年，SiO_2经时介质击穿（TDDB）的机制一直是一个非常活跃的研究领域，但是尚未得到一个毫无疑义的结论。随着工作电压和栅氧厚度的缩小，新的失效模式如软击穿也变得重要了。最近几年，一个重要的主题就是建立一个通过高压和高温加速寿命实验预测低压工作寿命的正确流程。一个相关的问题是：对于 SiO_2栅介质缩小的可靠性限制是什么？一些研究者相信随着微观和宏观栅氧均匀性的改善，SiO_2栅介质可以缩小到 20Å 以下，这与最近的制造实践结果一致。Weir 等人的一项研究结果证实，1.2V、70℃ 时 16Å 的氧化层是可能的[78]。基于目前的认识和 ITRS 提出的栅泄漏要求，SiO_2的

缩小可能被直接隧穿泄漏限制，而不是 TDDB 可靠性。

关于高 k 介质的 TDDB 可靠性的研究远少于关于 SiO_2的研究。一般地，许多高 k 介质都有更小的禁带宽度与更高的界面和体陷阱密度，这对它们的 TDDB 可靠性有不好的影响。现在已经有一些基于氟的栅介质的研究，这种介质被认为是最有前途的栅介质候选者。一些技术，如衬底和高 k 介质的氮化，表现出可以改善 HfO_2栅介质的击穿电压[79]。然而，高 k 介质其他不希望的效应，如迁移率衰退，是影响它被采用的主要因素。

热载流子效应　MOSFET 在正常工作时，高的横向电场会在漏附近产生热载流子，这将损伤栅介质并引起器件特性的衰退，如 g_m衰退和 V_{th}漂移。随着沟道长度的减小，对于给定的电源电压，沟道中峰值电场将增加。电源电压的降低将减轻这个问题。调整掺杂分布也可以改善热载流子可靠性。在现代 CMOS 技术下，电源电压可以减小到接近于 1V，因此热载流子的能量将小于热发射和碰撞电离所需的最小能量。然而，在电源电压减小到这些关键能量水平以下时，并没有观察到载流子衰退的突然变化[80]。另外一些热载流子衰退机制被提出以解释这些实验观察结果。虽然在实际应用中，热载流子不是工艺缩小的阻碍，但是热载流子可靠性将仍然是需要给以适当关注的议题。

负偏压温度不稳定性　负偏压温度不稳定性（NBTI）是 p-MOSFET 中阈值电压漂移的重要原因。在 NBTI 应力下，源、漏和衬底接地，而栅被偏置到一个负的应力电压。在高温时，这样一个偏置将增加栅氧中正电荷的产生和界面态的产生。因此，导致 g_m衰退和 V_{th}漂移。在一些模拟和混合信号应用中，需要阈值电压极度稳定。NBTI 是这类电路寿命的限制因素。Reddy 等人的研究报告报道了 NBTI 对现代数字电路的影响[81]，结果表明，静态 CMOS 工作时主要的衰退模式是 p-MOSFET NBTI。当工作频率下降时，受 NBTI 应力作用，环形振荡器的相对频率衰退增加。另外还发现，NBTI 应力导致了 SRAM 静态噪声裕量退化。可以预计，随着工艺的缩小，NBTI 对电路的影响日益重要。

最近一项研究显示，按照 V_{th}的漂移量达到一定值来定义寿命时，采用超薄氮化栅氧的 p-MOSFET 的寿命将主要由 NBTI 限制。在现代 CMOS 技术中，广泛使用氮化栅氧技术抑制硼穿通。然而，观察表明，氮化将减小 p-MOSFET 的 NBTI 寿命。在 Jeon 等人的研究中[82]，使用氮化栅氧厚度为 13Å 的 0.13μm CMOS 工艺制造 MOSFET。图 2-22 所示为在 NBTI 应力、HCI 应力下的 p-MOS 和在 HCI 应力下的 n-MOS 的寿命。由图可见，对 p-MOS，NBTI 应力下的寿命比 HCI 应力下的寿命低几个数量级，因此 NBTI 成为电路寿命的限制因素。将应力温度从 110℃升到 150℃，NBTI 寿命将更短。对于一些待选的高 k 介质，仅有非常有限的数据可以使用，并且对于电路级的广泛研究也很缺乏。可以看到，对当前和未来的 CMOS 技术，NBTI 可靠性对于器件技术和电路设计都是非常重要的议题。

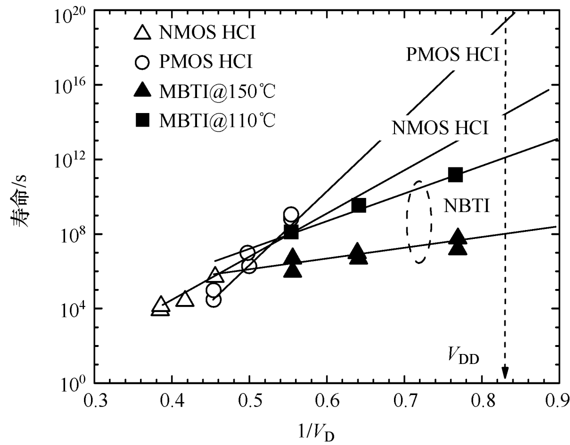


图 2-22　热载流子应力寿命和 0.13μm CMOS 技术中 PMOS NBTI 应力寿命
（在正常的 V_{dd} 值下，n-MOS 和 p-MOS 的 HCI 寿命远长于 NBTI 应力下 PMOS 的寿命）
（来源于参考文献［82］）

2.3　后道工序互连线技术

在 350nm 工艺之前，互连技术的发展在不断改善集成电路密度、性能和每项功能成本方面起着关键的作用。受性能改善的驱使，互连线缩小的连续进展导致了线条更厚、更窄、线间距减小、单位长度阻抗更低和电路工作频率更快。多层互连及低 k 介质提供了比先前铝互连线工艺更快的电路速度、减小的信号干扰和更好的可靠性等优势。然而，后道工艺（BEOL）的关注仍然是电迁移、热特性、工艺控制、多层整合和在未来片上系统（SoC）中最重要的高速全局信号等几个关键问题。在纳米时代，互连线时延已经是总时延的主要部分。如图2-23a

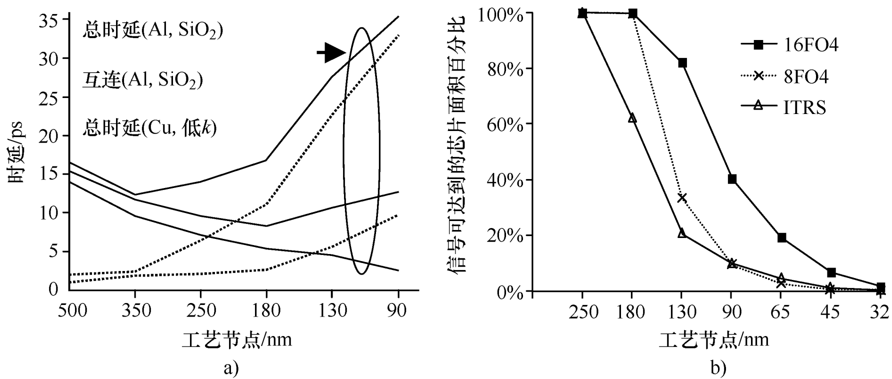


图 2-23　工艺缩小对芯片性能的影响：（总时延中占主要成分的互连线时延）
a）铝与铜中的互连线和栅时延　b）一个时钟周期信号可达到的面积百分比

所示，在 130nm 及更小时，即使是采用铜互连后道工艺，互连线时延会超过门时延。对 90nm 工艺，互连时延大约占总时延的 75%[1]。而且，由于芯片功能的不断增加，单个芯片的面积随着工艺的减小而增大，因而如图 2-23b[83] 中所示，在一个周期内将信号送到芯片的各个部分变得越来越困难。为实现频率提升，全局性信号需要采用流水线传递，当信号通路跨越功能模块时，将导致较高的等待时间和功率消耗。由于这些缩放效应，所有电路性能优化的重点就要从门级逻辑设计变成以互连为中心的设计。

2.3.1　互连缩放

与 FEOL 缩放能提高电路性能相比，由于互连线横截面面积的快速减小，互连尺寸的缩小却会产生较长的时延。为了应对 *RC* 时延的增加，从几何结构和材料方面，已经采用了各种的改进。如图 2-24 所示，典型多层结构的特性包括：

● 大纵横比（纵横比定义为互连的厚度与宽度之比）。因为金属到地的电容（C_g）约与线宽成比例，而金属阻抗（*R*）与横截面面积（线宽与线厚度的乘积）有关。当互连节距减小时，较大的纵横比可以实现较小的 *RC* 时延。这个趋势在全局互连中极受欢迎，而在局部互连时，较低的纵横比也是可以接受的，因为局部互连的连线较短，信号的路径时延对 *RC* 时延的增加不敏感。与高纵横比相关的两个问题需要加以注意。首先，要用金属完全填充一个深而窄的槽是很困难的。因此，纵横比大于 4 的互连和通孔的形成变得不可靠，特别是对于双镶嵌结构工艺。第二，对于较大的纵横比，互连线厚度的增加会导致与相邻互连之间产生较大的耦合电容（C_c），这会增加 *RC* 时延成分并且增加信号耦合噪声。这两个不希望的影响限制了未来技术换代中的这种缩放。

● 分层互连。在多层金属化中，缩减局部互连节距并增加其他互连层次中的厚度和宽度，可以满足更高的电路密度、减小的 *RC* 时延和更小的阻性损耗要求[83]。在过去的 10 年，可用的金属互连层数增加了 3 倍[84]，目前工艺（约90～180nm）中有 6～10 层金属，图 2-24 为一个多层互连实例示意图。局部范围和中等范围互连一般用于连接功能块内部的门，而全局互连一般用于实现不同功能单元之间的互连。除了信号传输（如数据

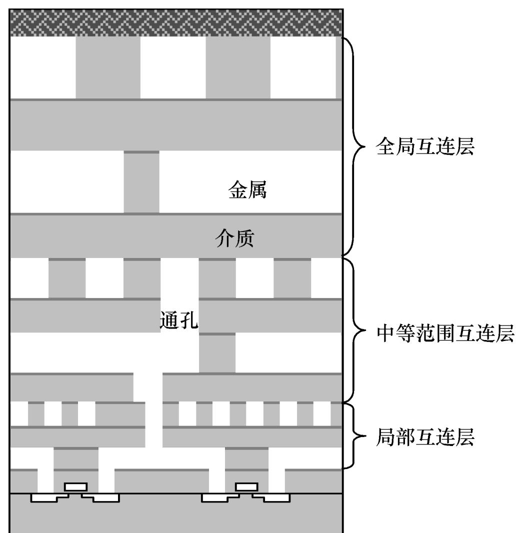


图 2-24　90nm 时多层互连结构的典型截面图

总线)，宽而厚的全局互连也用于时钟和电源线，或构成片上螺旋电感。同一层上的金属互连一般是平行分布，而与相邻层次上的互连线走向垂直。最近，对于全局性互连，对角线走向互连的引入比传统的直角互连结构的路径时延减小了19.8%，并且线长减小29.3%[85]。然而，在以前，这个技术对数据库的大小、光学邻近修正和掩膜写时间都提出了很大的挑战。这些问题的解决是未来技术成为可能的关键；电子设计自动化（EDA）公司和掩膜制造公司正在努力工作，以实现对角线走向互连。

● 局部和全局互连的不同缩小方式。为了与衬底上 MOSFET 的密度相匹配并维持 *RC* 时延，局部互连的节距与长度的缩小速度比垂直尺寸快得多。全局互连的缩小是由芯片边缘的长度决定的，其长度会随着栅尺寸的增大而增加。因此，全局互连的信号时延随着一代工艺到下一代工艺而增加，限制了整个芯片性能的提高（见图 2-23）。为了支持高性能处理器性能改善要求的频率提升，需要采用补偿技术，如增加转发器和流水线全局信号。

● 铜互连和低 *k* 介质。因为互连缩小是制约电路速度提高的主要瓶颈，业界做了很多的工作努力把新的 BEOL 材料整合到硅工艺中。两项主要的改进包括将金属材料铝改为铜、采用低 *k* 介质取代二氧化硅层。这些方式不仅有效地减小了单位长度互连的 *RC* 寄生参数，而且有助于改善 BEOL 的电迁移可靠性。另一方面，对互连最困难的挑战包括工艺集成、互连尺寸控制和电气可靠性。当工艺进入纳米时代，即使是低 *k* 铜，对于希望未来全局信号更快的要求来说也是不够的[85]。为了持续这种成功，新的 BEOL 技术包括材料和结构都是必需的。

2.3.2 铜互连技术

使用双镶嵌结构的铜互连制造是 1997 年由 IBM 引入的[86]。该技术的主要性能优点是铜的电阻率（1.8μΩ·cm，或者工艺完成后有效值为 2.2μΩ·cm）大约比铝的电阻率（3.3μΩ·cm）低 40%。因此对于同样的横截面，铜互连的 *RC* 时延大约比铝互连低 30% ~40%。铜互连的其他优点还包括由于铜原子具有较大的质量而改善了电迁移可靠性，以及由于一些工艺步骤的简化而降低了成本：因为在铝工艺中使用的干法腐蚀对刻蚀铜互连图形无效，作为替代，开发出了一种基于化学机械抛光（CMP）的镶嵌工艺。总的来说，随着低 *k* 介质（$k<3$）的使用、从 Al/SiO_2到铜/低 *k* 介质的转变使更高的互连密度、较少的金属层次、更快的信号传输和允许更大的电流密度成为可能，满足了当代超大规模集成电路系统高性能和高密度的要求。

使用铜技术除了带来优点，对于成功的纳米规模集成，从材料和工艺的角度来说，还存在严重的挑战。从材料方面来看，有两个主要的问题：材料电阻率和电迁移。当金属互连的截面尺寸下降到 sub-100nm 范围，电子在铜中的输运面临几个基本的物理限制。图 2-25 对作为线宽函数的铜线电阻率做了理论性预

测[87]。在 100nm 时，线宽接近铜晶粒尺寸，形态不完整（如表面粗糙散射）和晶粒边界散射将使铜的电阻率增加 100%（注意温度每升高 10℃ 电阻率将升高约 3.6%）。这些效应可以通过严格地控制铜的生长或生成单晶铜得到减缓。然而，在 100nm 工艺中，即使在单晶化后，由于铜被包在一个很窄的槽中，从而限制了晶粒尺寸和电阻率的改善。而且，在 45nm 工艺中，当线条尺寸接近铜中电子的平均自由程（39.3nm）时，金属表面散射将最终阻碍电子的传导。在这种环境下，铜电阻率的剧烈增加也必然导致 *RC* 信号时延和电源互连线上 *IR* 压降的极大增加。

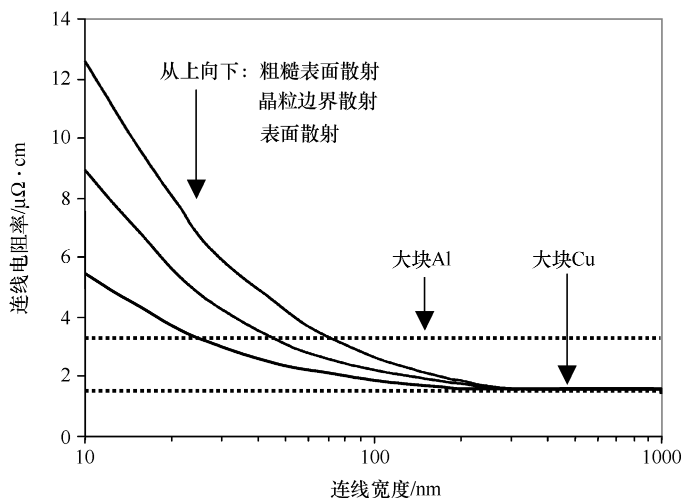


图 2-25　单个物理效应导致铜电阻率的增加（铜和铝的体电阻率作为参考[88]）

对于片上金属互连，要考虑的其他重要的材料属性是抗电迁移的能力。当互连上流过很大的电流密度时，金属原子将在电流的作用下沿着互连运动，这样将会在下游形成高浓度的积累，而在互连上留下不希望的空洞。由于铜的原子质量（63）大于铝（27），铜线本质上对这些不希望的电迁移效应有着更好的鲁棒性。依赖于测试条件和不同的金属层次，关键的失效电流密度 J_{crit}，对于铜线来说是铝线的两倍[8]。实验数据显示，在同样的电流密度下，铜的电迁移寿命是铝的 100 倍[86]。除了较高的 J_{crit} 值，铜较高的熔化温度（1034℃，而铝是 660℃）也对静电放电（ESD）可靠性有益处。另一方面，随着 MOSFET 密度和电路性能的改进，片上互连必须在更窄的互连传递日益增加的开关电流，互连抗电迁移的能力对器件的鲁棒工作很重要，因此当前铜工艺仍需要改善，特别是控制金属界面、晶粒尺寸和杂质[89,90]。

通常使用镶嵌工艺实现的铜互连涉及多道后工序，包括互连图形的形成、槽金属填充和使用 CMP 实现金属平坦化。尽管先进的技术要求对前道和后道工艺（FEOL 和 BEOL）中较小的特征尺寸都进行精确的光刻，而对于铜互连，槽填充和 CMP 控制还有独特的要求。例如，在镶嵌工艺中，首先在绝缘层中刻出槽，

然后通过电镀均匀地填充金属导体。通孔的形成采用相似的工艺。一个很重要的问题是铜的淀积质量，应该避免引起夹断和产生空洞，这两个不希望的现象将导致电阻增大和可靠性退化。虽然 IBM 已经实现了 0.2μm 宽、4∶1 纵横比互连线的无空洞填充，但是线宽的快速缩小和纵横比的增加要求一个更鲁棒的电镀工艺[86,91]。

在铜淀积步骤之后，金属和介质的表面使用化学机械技术抛光。虽然 CMP 工艺有较高的铜去除率，由于金属和介质的材料属性的不同，这种抛光是金属侵蚀和凹陷的来源。图 2-26 所示为 2μm 铜线在 CMP 步骤后的横截面 SE 微观图；研究者已经提出了一个描述金属侵蚀和凹陷的物理模型[92,93]。因为铜比介质软，它对化学浆料更加敏感，因此它的抛光速率更快。所以金属厚度比预期的要薄（侵蚀效应）并且在抛光后它的表面呈半球形（凹陷效应）。由于侵蚀，较高金属密度的区域将会有更大的损失，但是这个效应可以通过采用哑元金属图形的填充得到有效减小，控制金属密度的均匀性。然而，接地金属的填充增加了总的寄生电容，而悬空金属填充增加线间的容性耦合[92]。凹陷效应与金属的宽度有关，对较宽的金属，凹陷更严重。为了减小凹陷效应，对于较宽的互连，可以采用开槽，或者开孔（乳酪状孔）的方式以提高其有效硬度。采用开槽和开孔技术时，通常这些孔沿着电流流向排列以使电流集边效应最小，这将可以获得更好的电阻率。这在较宽的时钟线和电源分布网络中特别受欢迎。另外一种减小金属凹陷效应的方法是将较宽的互连分割成几条较窄的互连。然而，对于相同的有效电阻或电迁移限制，这个技术将消耗更多的面积。此外，开槽和开孔步骤要求对数据库进行后处理，这就要求指导工艺的设计规则必须考虑导致的电迁移极限和互连电阻率的衰退。研究显示，对于全局互连来说，考虑总面积开销和信号时延，互连分割的最优范围是 2 ~ 4 倍（即分割后，线宽的最优值是 3 ~ 5μm）[93]。注意这些方式并不单单在工艺控制中使用，在设计时也应予以考虑，特别是对高频信号趋肤深度小于互连宽度的情况。

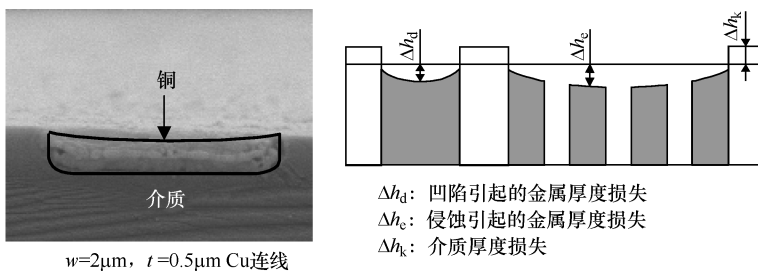


图 2-26　CMP 处理后，金属凹陷和腐蚀效应的 SEM 横截面图与模型

2.3.3　低 *k* 介质的挑战

为了满足减小互连寄生电容的要求，在 130nm 工艺中，多层互连结构中采

用了低介电常数（k）绝缘层，因此对于给定的金属间距，RC 时延和信号耦合都将减小。虽然作为绝缘层的 SiO_2 有许多天然的优点，但是它较高的相对介电常数（3.9）将导致较高的功耗和全局互连的 RC 时延。此外，对于全局互连，为了满足时序的约束，较高的 RC 时延要求更多的转发器，这将导致更大的芯片面积和功耗。为了减轻这些问题并支持数 GHz 级别的微处理器设计，需要采用新型的低 k 材料。在研究低 k 介质方面，目前，业界已经做出了许多切实的努力，已获得的低 k 材料包括，氟硅玻璃（$k=3.5$）、碳氧硅（$k=2.9$）、SiLK㊀（$k=2.7$）、聚对二甲苯（$k=2.3$）、Dendriglass（$k=1.5$）和干凝胶（$k=1.3$）。低 k 介质的终极目标是寻找到一种介电常数接近真空（$k=1.0$）的材料。最近，已经证明最小间距的互连线之间存有气隙，实际上这是在介质中有意生成的空洞。在以前的工艺中，出现这类空洞是对器件可靠性的威胁，但是现在已成为一种减小有效介电常数的方式。

因此，低 k 介质的引入对工艺集成提出了极大的挑战：必须在介电常数和可靠性之间做折衷。因为低 k 值介质主要是多孔的绝缘层（聚合体或硅化合物），其机械应力和热属性一般与金属互连不同。由于多孔结构的本身性质，低 k 材料的刻蚀速率很难控制，导致较差的侧墙覆盖[94]。另外，多孔低 k 介质的低热传导性引起很严重的焦耳热效应，在金属铜和绝缘层之间的界面引起热和机械应力失配[95]。并减小互连的电迁移性能退化。这些问题使表面粘附剧烈衰退并增加铜穿通进入介质的敏感性。因此，在实验温度下，铜/低 k 的平均寿命远低于铜/氧化层的寿命[96]。除了机械可靠性问题外，在金属间距接近 100nm 时，电可靠性，特别是经时介质击穿（TDDB），也已表现为一个重要的问题。虽然 TDDB 性能随着介电常数的减小而衰退，但是失效机制与多孔性的水平和热应力并没有关系，而主要与电场下铜离子的漂移有关[97,98]。

总之，虽然低 k 介质的应用改善了电路的性能，但是为了将介电常数 $k=2.1$ 的 2007 种材料集成起来，还需要解决热和电可靠性方面的阻碍[85]。

2.3.4 未来的全局互连技术

互连的基本目的是提供“在相距较远的不同点之间实现最小时延的通信”[99]。在纳米时代，当晶体管持续缩小以实现更小的特征尺寸和更快的工作时，全局互连成为提高信号速度和带宽的瓶颈。为了解决这个问题，各种各样的技术被引入到多层结构中，包括铜/低 k 材料、反向缩放和片上互连网络。材料、工艺、物理和成本本身存在的限制已经表明，当前采用的这些方式在 100nm 之

㊀ SiLK 是一种芳香族热固性有机材料，其成分处于保密状态——译者注。

下已不能满足要求[99,100]。

为了使未来高性能片上系统（SoC）集成成为可能，人们正在积极研究能够使速度更快、更可靠和成本更低的新颖全局互连技术。特别的例子包括系统级封装（SiP）、射频（RF）和光学互连以及三维集成。SiP 将全局互连和封装技术结合在一起，克服了 SoC 集成中的成品率退化问题[101]。由于在一个封装芯片中，是以相似的多层结构连接着不同的芯片和模块，与片上互连情况类似，因此开发 SiP 的设计原理相对容易。然而信号传输的基本限制，如渡越时间，仍然存在。射频（无线）或者光学互连改善了这个限制并通过使用片上无线网络或光学器件使之接近光速。考虑工艺的兼容性、设计开销和性能要求，对 45nm 工艺，当全局互连大于 32mm 时，它们将很受欢迎[102]。通过堆叠多个芯片在一起，三维集成缩短了最长的全局互连从而减小信号延迟。然而三维系统的散热和输入-输出互连仍然是巨大的挑战[99]。随着纳米时代互连技术的持续改进，先进的纳米技术，如纳米炭管和分子晶体互连，将给未来的计算系统提供更先进的互连。

参 考 文 献

[1] A. Allan et al., International Technology Roadmap for Semiconductors, *http://public.itrs.net*.

[2] R. Bez et al., Introduction to flash memory, *Proc. IEEE*, Vol. 91, No. 4, 2003.

[3] S. Thompson et al., A 90 nm logic technology featuring 50 nm strained silicon channel transistors, 7 layers of Cu interconnects, low κ ILD, and 1 μm^2 SRAM cell, *IEEE International Electron Devices Meeting*, 2002.

[4] P. A. Kraus et al., Scaling plasma nitrided gate dielectrics to the 65 nm node, *Semiconductor Fabtech*, 19th ed., FT 19-13/1 2003.

[5] M. L. Green et al., Ultrathin (<4 nm) SiO_2 and Si–O–N gate dielectric layers for silicon microelectronics: understanding the processing, structure, and physical and electrical limits, *J. Appl. Phys.*, Vol. 90, No. 5, 2001.

[6] G. D. Wilk, R. M. Wallace, and J. M. Anthony, High-k gate dielectrics: current status and materials properties considerations, *J. Appl. Phys.*, Vol. 89, No. 10, 2001.

[7] S. M. Sze, *Physics of Semiconductor Devices*, Wiley-Interscience, New York, 1981.

[8] D. K. Schroder, *Semiconductor Material and Device Characterization*, Wiley, New York, 1990.

[9] K. F. Schuegraf, C. C. King, and C. Hu, Ultra-thin silicon dioxide leakage current and scaling limit, *Digest of Technical Papers, Symposium on VLSI Technology*, 1992.

[10] T. Hori, *Gate Dielectrics and MOS ULSIs*, Springer-Verlag, New York, 1997.

[11] W. C. Lin and C. Hu, Modeling CMOS tunneling current through ultra thin gate oxide due to conduction- and valence-band electron and hole tunneling, *IEEE Trans. Electron Devices*, Vol. 48, No. 7, 2001.

[12] E. Vogel, Measurement of equivalent oxide thickness, *ITRS Document*, 2003.

[13] K. Ahmed et al., Impact of tunnel currents and channel resistance on the characterization of channel inversion layer charge and polysilicon-gate depletion of sub-20A gate oxide MOSFET's, *IEEE Trans. Electron Devices*, Vol. 46, No. 8, 1999.

[14] S. H. Lo et al., Modeling and characterization of n+- and p+-polysilicon-gated ultra thin oxides (21-26A), *Digest of Technical Papers, Symposium on VLSI Technology*, 1997.

[15] S. H. Lo et al., Quantum-mechanical modeling of electron tunneling current from the inversion layer of ultra-thin-oxide nMOSFET's, *IEEE Electron Dev. Lett.*, Vol. 18, No. 5, 1997.

[16] C. A. Richter et al., A comparison of quantum-mechanical capacitance voltage simulators, *IEEE Electron Device Lett.*, Vol. 22, No. 1, 2001.

[17] J. R. Hauser and K. Ahmed, Characterization of ultra-thin oxides using electrical $C-V$ and $I-V$ measurements, *Characterization and Metrology for ULSI Technology International Conference*, American Institute of Physics, 1998.

[18] G. Lucovsky et al., Separate and independent reductions in direct tunneling in oxide/nitride stacks with monolayer interface nitridation associated with (i) interface nitridation and (ii) increased physical thickness, *J. Vac. Sci. Technol. A*, Vol. 18, No. 4, 2000.

[19] Xin Guo and T. P. Ma, Tunneling leakage current in oxynitride: dependence on oxygen/nitrogen content, *IEEE Electron Device Lett.*, Vol. 19, No. 6, 1998.

[20] P. A. Kraus et al., Fundamental limits of MOSFET scaling with Si–O–N gate dielectrics, *Applied Materials ET Conference*, 2003.

[21] Y. C. Yeo, T. J. King, and C. Hu, MOSFET gate leakage modeling and selection guide for alternative gate dielectrics based on leakage considerations, *IEEE Trans. Electron Devices*, Vol. 50, No. 4, 2003.

[22] P. A. Kraus, K. Ahmed, T. C. Chua, M. Ershov, H. Karbasi, C. S. Olsen, F. Nouri, J. Holland, R. Zhao, G. Miner, and A. Lepert, Low-energy Nitrogen Plasmas for 65-nm node oxynitride gate dielectrics: a correlation of plasma characteristics and device parameters, *Digest of Technical Papers, Symposium on VLSI Technology*, 2003.

[23] E. M. Vogel, W. L. Hill, V. Misra, P. K. McLarty, J. J. Wortman, J. R. Hauser, P. Morfouli, G. Ghibaudo, and T. Ouisse, Mobility behavior of n-channel and p-channel MOSFET's with oxynitride gate Dielectric formed by low-pressure rapid thermal chemical vapor deposition, *IEEE Trans. Electron Devices*, Vol. 43, No. 5, 1996.

[24] A. Hegedus, C. Olsen, N. Kuan, and J. Madok, Clustering of plasma nitridation and post anneal steps to improve threshold voltage repeatability, *IEEE Trans. Semicond. Manuf.*, Vol. 16, No. 2, 2003.

[25] S. Inumiya et al., *Digest of Technical Papers, Symposium on VLSI Technology*, 2003.

[26] C. Hobbs et al., Fermi level pinning at the polySi/metal oxide interface, *Digest of Technical Papers, Symposium on VLSI Technology*, 2003.

[27] E. C. Carr, Role of interfacial nitrogen in improving thin silicon oxides grown in N_2O, *Appl. Phys. Lett.*, Vol. 63, No. 5, 1993.

[28] H. Niimi et al., Ultra thin oxide gate dielectrics prepared by low temperature remote plasma-assisted oxidation, *Surf. Coat. Technol.*, Vol. 98, 1998.

[29] S. V. Hattangady et al., Integrated processing of silicon oxynitride films by combined plasma and rapid-thermal processing, *J. Vac. Sci. Technol. A*, Vol. 14, No. 6, 1996.

[30] H. Stohr and W. Klemm, Uber Zweistoffsystem mit Germanium, I. Germanium/Aluminium, Germanium/Zinn und Germanium/Silicon, *Z. Anorg. U. Allgem. Chem.*, Vol. 241, No. 305, 1939.

[31] C. C. Wang and B. H. Alexander, Energy gap of germanium-Silicon alloys, *American Institute of Mining and Metallurgy Engineering Symposium*, 1954.

[32] E. R. Johnson and S. M. Christian, Some properties of germanium–silicon alloys, *Phys. Rev.*, Vol. 95, No. 560, 1954.

[33] R. Braunstein, R. Moore, and F. Herman, Intrinsic optical absorption in germanium-silicon alloys, *Phys. Rev.*, Vol. 109, No. 695, 1958.

[34] R. Braunstein, Lattice vibration spectra of germanium-silicon alloys, *Phys. Rev.*, Vol. 130, 1963.

[35] R. Braunstein and E. O. Kane, The valence band structure of the III-V compounds, *J. Phys. Chem. Solids*, Vol. 23, 1962.

[36] R. Braunstein, Valence band structure of germanium-silicon alloys, *Phys. Rev.*, Vol. 130, 1954.

[37] F. Herman, Speculation on the energy band structure of Ge–Si alloys, *Phys. Rev.*, Vol. 95, 1954.

[38] R. People, Indirect band gap of coherently strained Ge_xSi_{1-x} bulk alloys on <001> silicon substrates, *Phys. Rev. B*, Vol. 32, No. 2, 1985.

[39] R. People, Physics and applications of Ge_xSi_{1-x} strained-layer heterostructures, *IEEE J. Quantum Electron.*, Vol. 22, No. 9, 1986.

[40] C. G. Van de Walle, Theoretical study of Si/Ge interfaces, *J. Vac. Sci. Technol. B*, Vol. 3, No. 4, 1985.

[41] C. G. Van de Walle, Theoretical calculations of heterojunction discontinuities in the Si/Ge system, *Phys. Rev. B*, Vol. 34, No. 8, 1986.

[42] A. G. O'Neill and D. A. Antoniadis, Deep submicron CMOS based on silicon germanium technology, *IEEE Trans. Electron Devices*, Vol. 43, 1996.

[43] K. L. Wang, S. G. Thomas, and M. O. Tanner, SiGe band engineering for MOS, CMOS and quantum effect devices, *J. Mater. Sci.*, Vol. 6, 1995.

[44] M. C. Ozturk, N. Pesovic, I. Kang, J. Liu, H. Mo, and S. Gannavaram, Ultra-shallow source/drain junctions for nanoscale CMOS using selective silicon–germanium technology, *Jpn. Soc. Appl. Phys. IWJT*, 2001.

[45] M. C. Ozturk, J. Liu, H. Mo, and N. Pesovic, Advanced $Si_{1-x}Ge_x$ source/drain and contact technologies for sub-70 nm CMOS, *IEEE*, 2002.

[46] P. Ranade, H. Takeuchi, V. Subramanian, and T.-J. King, A novel elevated source/drain PMOSFET formed by Ge–B/Si intermixing, *Electron Device Lett.*, Vol. 23, No. 4, 2002.

[47] P. Ranade et al., Application of silicon–germanium in the fabrication of ultra-shallow extension junctions for sub-100 nm PMOSFETs, *IEEE Trans. Electron Devices*, Vol. 49, No. 8, 2002.

[48] J. M. Harmann et al., Effect of HCl on the SiGe growth kinetics and reduced pressure–chemical vapor deposition, *J. Cryst. Growth*, Vol. 241, 2002.

[49] R. B. Fair, *Rapid Thermal Processing: Science and Technology*, 1993.

[50] B. Ramachandran, H. Forstner, and E. Chiao, Beyond the 100 nm node: single-wafer RTP, *Solid State Technol.*, May 2003.

[51] B. Peuse, G. Miner, and M. Yam, Method and apparatus for measuring substrate temperatures, U.S. patent 5, 660, 472, 1997.

[52] C. Schietinger, B. Adams, and C. Yarling, Ripple technique: a novel non-contact wafer emissivity and temperature method for RTP, *Mater. Res. Symp. Proc.*, Vol. 224, No. 23, 1991.

[53] T. Ito et al., 14 nm-depth low resistance boron doped extension by optimized flash lamp annealing, *International Symposium on Semiconductor Manufacturing*, 2002.

[54] S. Talwar, Y. Wang, and M. Thompson, Laser annealing for junction fabrication in CMOS devices, *Electrochemical Society Spring Meeting*, 2003.

[55] J. Kittl, *VLSI Technology Symposium*, 1996.

[56] J. P. Lu et al., A novel nickel SALICIDE process technology for CMOS devices with sub-40 nm physical gate length, *International Electron Device Meeting*, 2002.

[57] J. Kittl et al., Silicides for 65 nm CMOS and beyond, *Electrochemical Society Meeting*, 2003.

[58] *http://public.itrs.net/.*

[59] S. Kuppurao, H. Joo, and G. Miner, In situ steam generation: a new rapid thermal oxidation technique, *Solid State Technol.*, 2000.

[60] K. Reid et al., Dilute steam rapid thermal oxidation for 30 Å gate oxides, *Electrochemical Society Meeting*, 1999.

[61] P. Meissner, A. Hegedus, J. Madok, R. Thakur, and G. Miner, Thermal technologies for sub-100 nm CMOS scaling: development strategies, *Electrochemical Society Meeting*, 2002.

[62] G. Timp et al., Low leakage, ultra-thin gate oxides for extremely high performance sub-100 nm nMOSFETs, *Technical Digest, IEEE International Electron Devices Meeting*, Vol. 930, 1997.

[63] K. Ahmed et al., Applied Materials internal document.

[64] B. Yu, H. Wang, A. Joshi, Q. Xiang, E. Ibok, and M.-R. Lin, 15 nm gate length planar CMOS transistor, *Technical Digest, IEEE International Electron Devices Meeting*, pp. 937–939, Dec. 2001.

[65] Y.-C. King, H. Fujioka, S. Kamohara, W.-C. Lee, and C. Hu, Ac charge centroid model for quantization of inversion layer in n-MOSFET, *Proceedings of the International Symposium on VLSI Technology, Systems and Applications*, pp. 245–249, June 1997.

[66] International Technology Roadmap for Semiconductors, *http://public.itrs.net.*

[67] Q. Lu, R. Lin, P. Ranade, Y.-C. Yeo, X. Meng, H. Takeuchi, T.-J. King, C. Hu, H. Luan, S. Lee, W. Bai, C.-H. Lee, D.-L. Kwong, X. Guo, X. Wang, and T.-P. Ma, Molybdenum metal gate MOS technology for post-SiO_2 gate dielectrics, *Technical Digest, IEEE International Electron Devices Meeting*, pp. 641–644, 2000.

[68] Y.-C. Yeo, P. Ranade, Q. Lu, R. Lin, T.-J. King, and C. Hu, Effects of high-κ dielectrics on the work-functions of metal and silicon gates, *Proceedings of the Symposium on VLSI Technology*, pp. 49–50, Kyoto, Japan, June 2001.

[69] L. Chang, S. Tang, T.-J. King, J. Bokor, and C. Hu, Gate length scaling and threshold voltage control of double-gate MOSFETs, *Technical Digest, IEEE International Electron Devices Meeting*, pp. 719–722, Dec. 2000.

[70] S. H. Lo, D. A. Buchanan, Y. Taur, and W. Wang, Quantum-mechanical modeling of electron tunneling current for the inversion layer of ultra-thin-oxide nMOSFETs, *IEEE Electron Device Lett.*, pp. 209–211, May 1997.

[71] W.-C. Lee and C. Hu, Modeling gate and substrate currents due to conduction- and valence band electron and hole tunneling, *Proceedings of the Symposium on VLSI Technology*, pp. 198–199, 2000.

[72] Y.-C. Yeo, Q. Lu, W.-C. Lee, T. King, C. Hu, X. Wang, X. Guo, and T. P. Ma, Direct tunneling gate leakage current in transistors with ultra-thin silicon nitride gate dielectric, *IEEE Electron Device Lett.*, Vol. 21, No. 11, pp. 540–542, Nov. 2000.

[73] C. Hu, A compact model for rapidly shrinking MOSFETs, *Technical Digest, IEEE International Electron Devices Meeting*, pp. 285–288, Dec. 2001.

[74] M. V. Dunga, X. Xi, J. He, I. Polishchuk, Q. Lu, M. Chan, A. M. Niknejad, and C. Hu, Modeling of direct tunneling current in multi-layer gate stacks, Workshop on Compact Modeling, *6th International Conference on Modeling and Simulation of Microsystems*, San Francisco, Feb. 2003.

[75] R. Shrivastava and K. Fitzpatrick, A simple model for the overlap capacitance of a VLSI MOS device, *IEEE Trans. Electron Devices*, Vol. 29, No. 12, Dec. 1982.

[76] S. Thompson, P. Packan, T. Chani, M. Stettler, M. Alavi, I. Post, S. Tyagi, S. Ahmed, S. Yang, and M. Bohr, Source/drain extension scaling for 0.1 μm and below channel length MOSFETs, *Proceedings of the Symposium on VLSI Technology*, pp. 132–133, Honolulu, HI, June 1998.

[77] B. Cheng, M. Cao, R. Rao, A. Inani, P. V. Voorde, W. M. Greene, J. M. C. Stork, Z. Yu, P. M. Zeitzoff, and J. C. S. Woo, The impact of high-κ gate dielectrics and metal gate electrodes on sub-100 nm MOSFETs, *IEEE Trans. Electron Devices*, Vol. 46, No. 7, pp. 1537–1542, July 1999.

[78] B. E. Weir, P. J. Silverman, M. A. Alam, F. Baumann, D. Monroe, A. Ghetti, J. D. Bude, G. L. Timp, A. Hamad, T. M. Oberdick, N. X. Zhao, Y. Ma, M. M. Brown, D. Hwang, T. W. Sorsch, and J. Madic, Gate oxides in 50 nm devices: thickness uniformity improves projected reliability, *IEEE International Electron Devices Meeting*, pp. 437–440, 1999.

[79] Q. Lu, H. Takeuchi, X. Meng, T.-J. King, C. Hu, K. Onishi, H.-J. Cho, and J. C. Lee, Improved performance of ultra-thin HfO_2 CMOSFETs using poly-SiGe gate, *Symposium on VLSI Technology*, pp. 86–87, June 2002.

[80] J. Chung, M.-C. Jeng, J. Moon, P. K. Ko, and C. Hu, Low voltage hot-electron currents and degradation in deep submicrometer MOSFETs, *IEEE Trans. Electron Devices*, Vol. 37, No. 7, pp. 1651–1657, July 1990.

[81] V. Reddy, A. T. Krishnan, A. Marshall, J. Rodriguez, S. Natarajan, T. Rost, and S. Krishnan, Impact of negative bias temperature instability on digital circuit reliability, *International Reliability Physics Symposium Proceedings*, pp. 248–254, Apr. 2002.

[82] C.-H. Jeon, S.-Y. Kim, H.-S. Kim, and C.-B. Rim, The impact of NBTI and HCI on deep sub-micron PMOSFETs' lifetime, *IEEE Integrated Reliability Workshop Final Report*, pp. 130–132, 2002.

[83] V. Agarwal, M. S. Hrishikesh, S. W. Keckler, and D. Burger, Clock rate versus IPC: the end of the road for conventional microarchitectures, *Proceedings of the International Symposium on Computer Architecture*, pp. 248–259, 2000.

[84] C. W. Kaanta et al., Submicron wiring technology with tungsten and planarization, *Proceedings of the IEEE International Electron Devices Meeting*, pp. 209–212, Dec. 1987.

[85] M. Igarashi et al., A diagonal-interconnect architecture and its application to RISC core design, *Digest of the IEEE International Solid-State Circuits Conference*, pp. 166–167, Feb. 2002.

[86] D. Edelstein et al., Full copper wiring in a sub-0.25 μm CMOS ULSI technology, *Proceedings of the IEEE International Electron Devices Meeting*, pp. 773–776, 1997.

[87] S. M. Rossnagel and H. Kim, From PVD to CVD to ALD for interconnects and related applications, *Proceedings of the 2001 International Interconnect Technology Conferences*, pp. 3–5, 2001.

[88] A. E. Kaloyeros, E. T. Eisenbraun, J. Welch, and R. E. Geer, Exploiting nanotechnology for terahertz interconnect, *Semicond. Int.*, pp. 56–59, Jan. 2003.

[89] S. Voldman, R. Gauthier, D. Reinhart, and K. Morrisseau, High-current transmission line pulse characterization of aluminum and copper interconnects for advanced CMOS semiconductor technologies, *IEEE 36th Annual International Reliability Physics Symposium*, pp. 293–301, 1998.

[90] C.-K. Hu and J. M. E. Harper, Copper interconnect: fabrication and reliability, *International Symposium on VLSI Technology, Systems, and Applications*, pp. 18–22, 1997.

[91] D. Chung, J. Korejwa, and E. Walton, Introduction of copper electroplating into a manufacturing fabricator, *IEEE/SEMI Advanced Semiconductor Manufacturing Conference*, pp. 282–289, 1999.

[92] B. E. Stine et al., The physical and electrical effects of metal-fill patterning practices for oxide chemical–mechanical polishing processes, *IEEE Trans. Electron Devices*, Vol. 45, No. 3, pp. 664–678, Mar. 1998.

[93] R. Chang, Y. Cao, and C. Spanos, Modeling metal dishing effect for interconnect process-design co-optimization, *IEEE International Electron Devices Meeting*, 2003.

[94] S. Purushothaman et al., Opportunities and challenges in ultra low κ dielectrics for interconnect applications, *IEEE International Electron Devices Meeting*, pp. 529–532, Dec. 2001.

[95] H. Wu, J. Cargo, C. Peridier, and J. Serpiello, Reliability issues and advanced failure analysis deprocessing techniques for copper/low-κ technology, *IEEE 41st Annual International Reliability Physics Symposium*, pp. 536–544, 2003.

[96] K.-D. Lee et al., Electromigration study of Cu/low κ dual-damascene interconnects, *IEEE 40th Annual International Reliability Physics Symposium*, pp. 322–326, 2002.

[97] J. Noguchi et al., Impact of low-κ dielectrics and barrier metals on TDDB lifetime of Cu interconnects, *IEEE 39th Annual International Reliability Physics Symposium*, pp. 355–359, 2001.

[98] E. T. Ogawa et al., Leakage, breakdown, and TDDB characteristics of porous low-κ silica-based interconnect dielectrics, *IEEE 41st Annual International Reliability Physics Symposium*, pp. 166–172, 2003.

[99] J. D. Meindl et al., Interconnecting device opportunities for gigascale integration (GSI), *IEEE International Electron Devices Meeting*, pp. 525–528, 2001.

[100] T. Ohba, Multilevel interconnect technologies in SoC and SiP for 100-nm node and beyond, *6th International Conference on Solid-State and Integrated-Circuit Technology*, pp. 46–51, Oct. 2001.

[101] K. L. Tai, System-in-package (SIP): challenges and opportunities, *IEEE Asia and South Pacific Design Automation Conference*, pp. 191–196, Jan. 2000.

[102] A. V. Mule et al., Towards a comparison between chip-level optical interconnections and board-level exterconnection, *IEEE International Interconnect Technology Conference*, pp. 92–94, 2002.

第 3 章　亚波长光刻的理论与实践

3.1　引言与成像理论概述

在竞争激烈的芯片制造领域，器件必须向更小更快的方向发展才能跟上器件集成度每两年翻番的规律。下面几章将深入讨论决定集成度能否翻番的一项关键工艺——光刻。我们将关注当前关于分辨率增强技术的理论与实现方法，包括曝光量的优化、相移掩膜版、光学邻近修正和亚分辨率辅助特性。贯穿本书的重点是设计和分辨率增强的相互影响。理解光学增强与设计之间的相互影响效应以及针对分辨率增强技术的设计优化需求，是 65nm 以及小于 65nm 节点技术成功实现光刻成像的关键。对于 100nm 和 65nm 节点，图形的形成仍是基于光学方法的光刻。当成像图形的特征尺寸小到与光刻采用的光的波长可以相比拟时，就不能采用光学系统实现这种小尺寸图形的成像。下述方程描述了特征尺寸与波长之间的关系：

$$d=\frac{k\lambda}{\mathrm{NA}}$$

式中，k 通常称为 k 因子，在光学系统中，能达到的最小 k 值是 $k=0.25$；λ 是所用光的波长；NA 是投影光学的数值孔径；d 是特征尺寸的半节距。

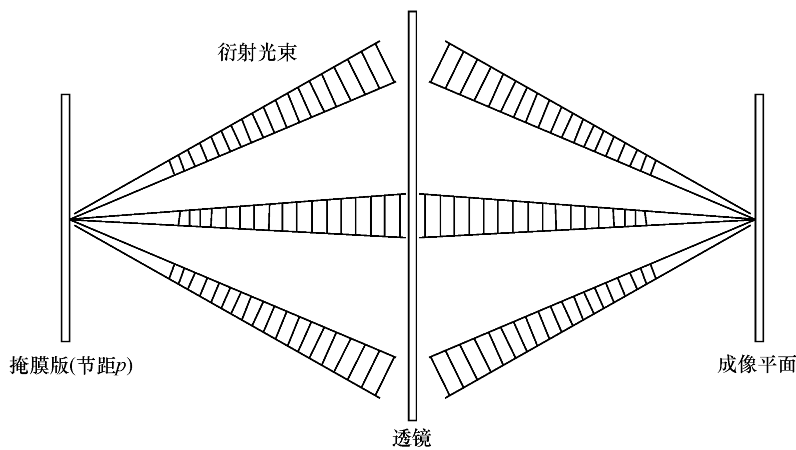


图 3-1　具有一定尺寸的掩膜图形成像的波动光学解释

（尺寸一定的图形产生了衍射光束，其衍射角由 Bragg 条件确定。衍射光束强度场分布是均匀的。透镜将衍射光束聚焦。聚焦通过相移实现，而该相移取决于到中心的距离。衍射光束的干涉重新形成了图像的调制。如果衍射光束没有穿过透镜，将不会形成成像）

图 3-1 所示为该方程的基本物理含义，在相干光照射下，节距 $p=2d$ 的掩膜图形将形成一系列衍射光束，其衍射角 α 满足方程 $n\lambda = p \cdot \sin\alpha$，式中 n 是任意整数。因此，随着节距的减小，衍射角将增大。对于这些衍射光，再通过具有确定数值孔径 NA 的透镜发生干涉，使发散的衍射光束重新聚焦成像。如果透镜的数值孔径太小，不能捕获对应于 $n = \pm 1$ 的衍射光束，则生成的像将没有空间调制。对图3-1所示情况，入射光垂直照射掩膜版，至少可以通过对应于 $n=0$，± 1 的三列光束成像。值得注意的是，即使透镜捕获了第一衍射级，生成的图像仍然会严重退化：即通过正方形掩膜图形生成的图像强度 I 呈现正弦分布，虽然还保留着空间间距，但是已不能明确显示所有的边界。

随着光刻图形特征尺寸越来越小，照射掩膜的将不再是简单的垂直入射光。控制投射到掩膜版上的入射光的角度分布是控制成像质量的光刻设备关键工艺参数之一。图 3-2 说明了常规照明的概念，具有一定入射角的入射光照射到掩膜版上。说明不同照明方式的简单办法是在一个平面上观察图像，与中心点的距离代表入射角，中心点对应于垂直入射。采用这种表示方法，常规照明看起来像一个圆，圆外径对应于锥形入射光的最大角。还有其他几种常用照明方式：环形、四极和偶极照明，也得到越来越广泛的使用。后面几章将对它们进行更详细的讨论。

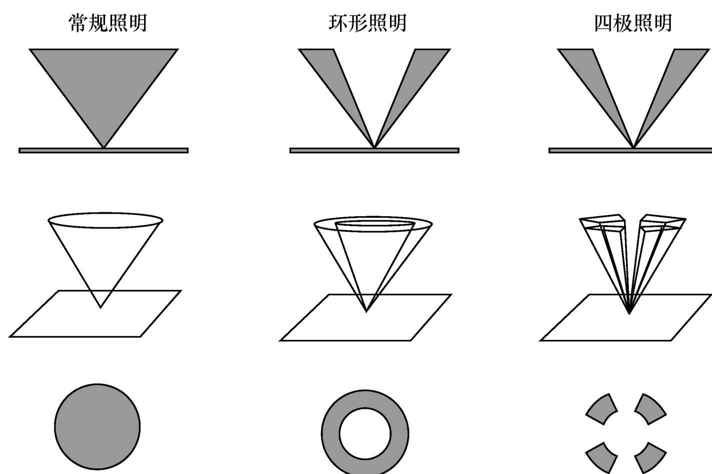


图 3-2　照明

（在光刻中使用的几种不同照明方式，图中显示了常规、环形和四极照明方式。图中描述了投射到掩膜版上光的角分布）

图 3-3 与图 3-4 所示为照明方式影响成像过程的机理。不同角度的光照射掩膜形成的衍射图形是每个角度衍射光的叠加。也就是说，对于如果只受到垂直入射时将生成如图 3-3b 所示的衍射图形的掩膜，在如图 3-3a 中所示的照明模式作用下，最后形成的衍射图形将如图 3-3c 所示。正如上面简单实例中讨论的，衍

射图形中落在系统数值孔径定义的圆的外面的那一部分，对于成像没有任何贡献。当掩膜图形尺寸缩小时，衍射级次发生的角度越来越大。这样，一级衍射光中将会有越来越大的比例不能穿过透镜，导致成像性能退化，如图 3-4 所示。

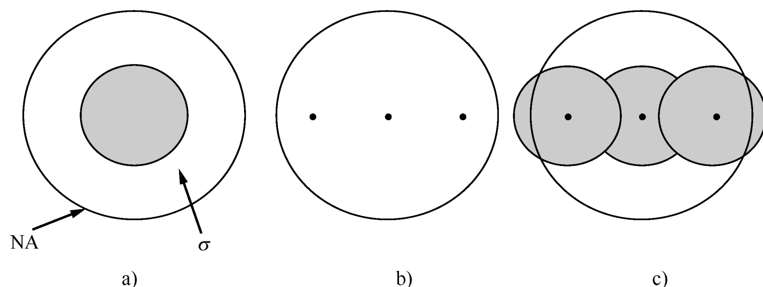


图 3-3　部分相干光情况下图像形成过程示意图

a）透镜和照明设置　b）掩膜衍射级次　c）衍射图形

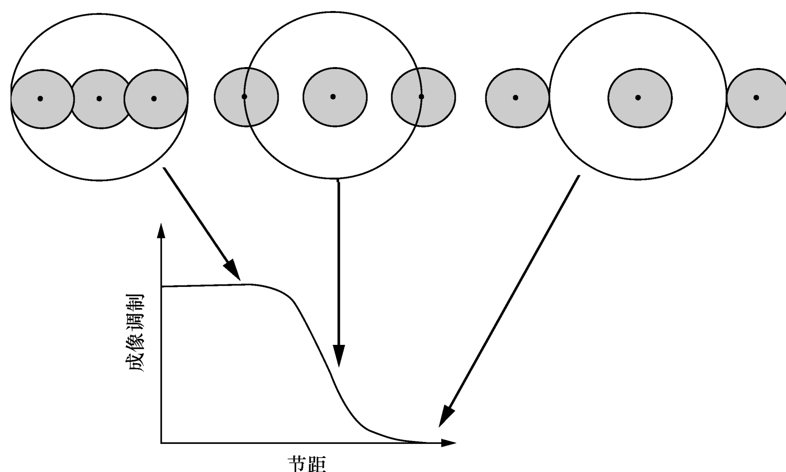


图3-4　相干对成像性能的影响：图像调制随着图形尺寸的减小而逐渐衰退

3.2　对 100nm 节点的挑战

100nm 节点光刻面临的挑战是与低 k 因子成像联系在一起的。k 因子接近 0.25 时，对于较宽的工艺波动范围，通过光刻获得满足要求的成像质量的能力下降了。因此，为了分析 100nm 节点的挑战，需要分析 100nm 节点工艺期望的 k 因子、在制造环境中存在的工艺波动类型以及低 k 因子是如何导致对工艺波动更加敏感。

3.2.1　100nm 节点的 k 因子

图 3-5 所示为当工艺节点的特征尺寸不断缩小时，光刻技术所面临的日益严

峻的问题。特征尺寸的减小速度已经超过了曝光波长的减小。从 1980 年到现在，批量生产和试制生产中使用的曝光波长已经从 436nm 下降到 365nm、248nm 直到 193nm，缩小了 243nm（约为原来的 44%）。而在这段期间，特征尺寸由 1.5μm 下降到 65nm，缩小了 1.435μm（约为原来的 4.3%），比波长的减小速率快了一个数量级。按照摩尔定律描述的特征尺寸的快速减小，很大程度上是由日益减小的图像尺寸与曝光波长之比来实现的。大约在 1995 年，光刻技术进入了特征尺寸小于成像光波波长的时代，这个时期被称作亚波长时代[1]。正如我们看到的，只采用波长并不是评估光刻能力的最好方式；由半节距与数值孔径乘积再除以波长所定义的 k 因子是更好的评估方式。不同时期技术节点对应的 k 因子变化趋势如图 3-6 所示。由图可见，很幸运的是光刻机的能力改善远快于曝光波长趋势图，这主要是由于透镜数值孔径提高的结果，光刻设备由较早的 0.28NA 提高到 0.85NA。虽然如此，对光刻的要求仍在剧烈地增长，并且 k 因子正快速接近最小可能值 $k=0.25$。对于 90nm 工艺技术，考虑到工艺将采用 NA 为 0.75 的 193nm 曝光机，预计 k 因子大约是 0.34。而对于 65nm 工艺，要求还不是很清晰，在 65nm 开始阶段，生产可能将依赖超高 NA（≥0.85）的 193nm 光刻

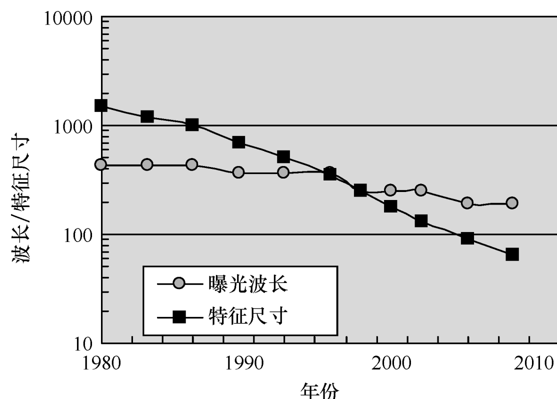


图 3-5 在半导体制造中曝光波长与最小特征尺寸演化情况对比

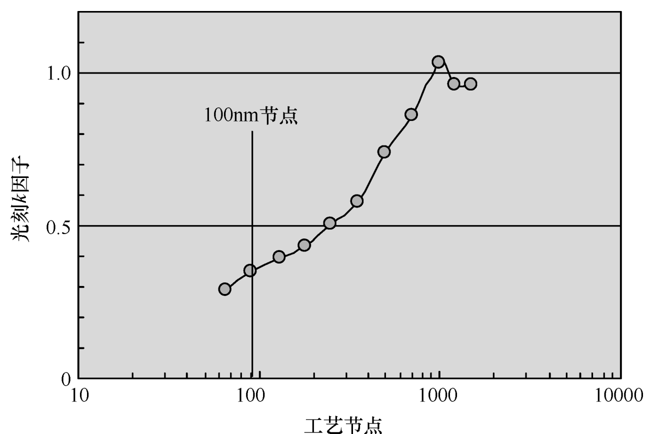


图 3-6 不同工艺节点 k 因子的变化情况

机，等效于 $k\approx0.29$。采用波长为 157nm 的下一代光刻机将可能是生产的主流选择。

3.2.2　重要的工艺波动

在分析低 k 因子成像技术需要应对的工艺波动之前，本节简单讨论在光刻工艺中遇到的一些工艺波动。工艺波动可以根据光刻工艺的关键元素分类，其中最重要的包括：晶圆材料、掩膜、照明系统、投影透镜、聚焦系统和光刻胶。下面简要分析波动类型，但决不是所有可能波动的完整罗列。

多种机制使晶圆成为光刻工艺中的波动源之一。例如，晶圆工业已经过前一步光刻形成的图像会导致反射率的局部变化。通常碰到的情况是，有些光刻胶线条将跨越先前已形成的金属层图形，这将使下面金属层图形附近的光刻胶线条反射率增加。反射率的调制将引起局部有效光强度的变化，这又会进而引起线宽的局部变化，并且在极端情况下引起灾难性的失效[2]。对于覆盖在硅有源区以及浅槽介质上的光刻胶，也会经常出现相似的情况。这些区域存在的不同材料影响着反射率，导致了不希望的线宽变化。虽然使用抗反射层（ARC）能减小反射率变化的影响[3]，但是这并不是 100% 的有效。在某些情况下，例如用于离子注入的光刻胶，使用抗反射层也有问题，因为在注入前必须去除抗反射层，而去除抗反射层要增加工艺步骤，导致成本太高，这都是不希望的。从空间像的观点看，较大的空间像光强 I 斜率 $\mathrm{d}I/\mathrm{d}x$，将使反射率变化的影响最小。在这种情况下，反射率导致的光强变化（$\mathrm{d}I$）对图像成像线宽偏移（$\mathrm{d}x$）的影响较小。

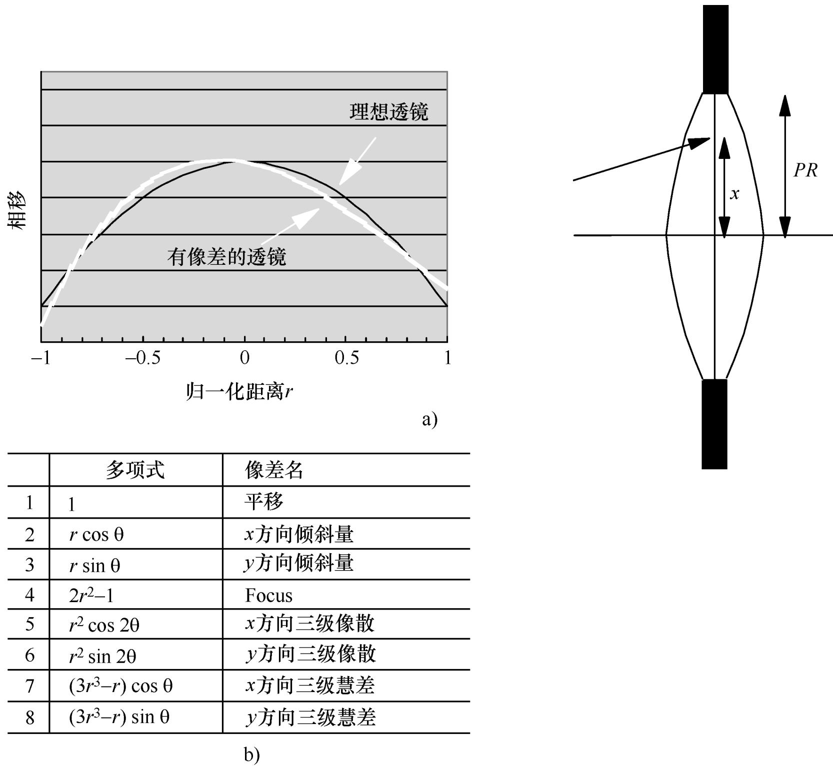
另一个由晶圆引入的相当普遍的工艺波动，是局部或整个晶圆范围的表面形貌变化。在较老的工艺中，局部表面形貌变化只是由金属层刻蚀后再淀积氧化层形成的，在金属区上形成小丘而在无金属区形成凹陷。现在这些台阶高度已变得不能接受，因此现在的工艺使用化学机械抛光（Chemical-Mechanical Polishing，CMP）使表面平坦。尽管 CMP 使工艺水平提高了一大步，但是它们的平坦化能力并不十分完美。例如，由于抛光速率与图形密度有关，就导致了整个晶圆范围以及局部区域都存在薄膜厚度的变化。因此，在芯片边缘以及图形密度与平均图形密度存在较大差别的区域，抛光后将可能还留下有残存的材料。

晶圆的局部表面形貌通过两种机制来影响成像。首先，它引起反射率的局部变化，这不仅是由透明材料的厚度变化引起，而且抗反射材料和光刻胶膜的厚度变化也会引起反射率的局部变化。第二，曝光机需要修正晶圆载物台与透镜之间的距离，使图像正好聚焦于晶圆顶层的表面。因为自动对焦系统仅能调整晶圆载物台的距离或者倾斜度，晶圆表面上比典型芯片尺寸（20mm × 20mm 的面积）小的范围内存在的局部高度调制将很难得到补偿。同样，在晶圆边缘普遍存在的晶圆弯曲也无法修正。焦距变化将导致特征尺寸的改变，在极端情况下会造成灾难性的图形失效。图像的特征尺寸与焦距的关系不是线性的，一般情况下是二次

方关系，并且变化的幅度与光刻工艺和图形尺寸均有关。对于焦距的小范围变化所引起的图像变化，光刻一般有足够的容限，但是如果焦距变化太大，二次方关系将很快导致曝光尺寸产生不能接受的变化。

非理想的照明系统也会导致工艺波动。作为光刻机一部分的照明灯将光投射到掩膜版上，需要保持成像场中的光强均匀。对 25mm 宽的典型缝隙，实现不大于 1% 的波动是一个巨大的挑战。非均匀照明灯引起图形视场中本应等价的不同部分发生局部光强改变。因此它们与反射率变化的影响是相似的，并且要求空间像斜率足够大以使它们的影响降到最小。在这些影响中较难探测和比较敏感的情况是不同角度的亮度均匀性，而不是总的光照度[4]。例如，如果从照明灯的一侧发出的光强强于另一侧，由焦距的变化导致的图形变形或者漂移最显著。

用来实现掩膜成像的投影光学系统也是工艺波动的来源之一。投影透镜的非理想性通常称为像差，这是一个极受关注的领域[5-9]。它们可能是由于各种透镜单元的非理想安装或者是由于透镜制造过程中的容差带来的形状偏差引起的。这些偏差可以被描绘成光瞳面内特性分布的相位误差。基于轴向对称和半径的变化可以将像差分类。图 3-7 所示为像差的简单实例。在该例中，相位误差的增加量



	多项式	像差名
1	1	平移
2	$r\cos\theta$	x方向倾斜量
3	$r\sin\theta$	y方向倾斜量
4	$2r^2-1$	Focus
5	$r^2\cos 2\theta$	x方向三级像散
6	$r^2\sin 2\theta$	y方向三级像散
7	$(3r^3-r)\cos\theta$	x方向三级慧差
8	$(3r^3-r)\sin\theta$	y方向三级慧差

b)

图 3-7 透镜像差、特征、类型和衍射图形之间的关系

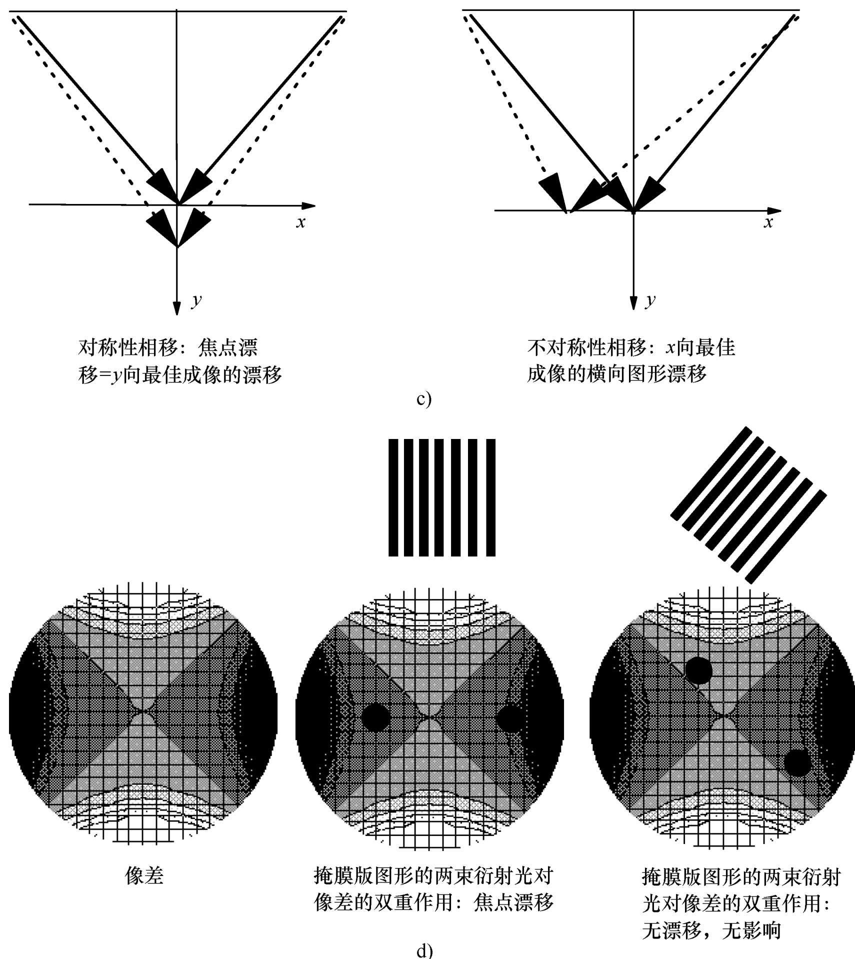


图 3-7　透镜像差、特征、类型和衍射图形之间的关系（续）

（图 a 所示为像差一般概念的描述。如 3.2 节中所描述的，透镜可以被当作像物镜。仅当透镜与衍射角的关系遵循理想的相移时才可以实现完美的性能。与完美的成像间的偏差被称为像差。归一化的距离是 x/PR，其中 PR 是指光瞳半径。从图 b 可以看到可能的相位偏差可以被描述为一组称之为 Zernike 多项式的基本函数的叠加。这些基本函数是归一化距离 r（取值范围由 0 ~ 1）以及光瞳面内角度 θ（0 ~ 360°）的函数，表格提供了典型像差的类型。图 c 给出了双光束干涉时像差对密集条状图形的影响：（左）两个级次的相移相等；（右）两个级次相移不同。图 d 给出了衍射图形和像差的关系（像散的实例）。图示的是第三级像散的相位误差。在光瞳面内图示的相位误差呈现鞍形形状。该图突出显示了像差对称性与图形对称性之间的关系）

与半径有关，并有双重对称，其中在左侧引入负的相位误差，而在右侧引入正的相位误差。图 3-7d 是一个简单掩膜图形（图形中多条线条走向相同）的衍射和像差的相位偏移之间相互影响的实例。当掩膜图形在垂直方向时，两条衍射光束

的相移量相等，等效于焦距的偏移，如图 3-7c 所示。若将同一掩膜图形旋转 90°，与垂直情况相比相移的幅度相同，但是符号相反。结果，由于不同方向图形对应的最佳焦距不同，产生的像差称作像散。如果掩膜旋转 45°，这个影响就会最小，因为在这种情况下衍射各级次和零级衍射没有相位漂移，因此像差对成像没有影响。这是一个像差依赖于成像图形类型、照明模式和图形方向的实例。通常像差会导致图形在光传播方向上（即在最好焦距上的漂移）的空间位移或在垂直于光轴方向的偏移。由于在透镜测量能力、透镜制造方面的重大改进和透镜组装技术的进步，在减小半导体制造中由透镜带来的像差大小方面已经取得了巨大进展。

从曝光时掩膜和晶圆保持静止的步进重复系统到掩膜和晶圆同时移动的步进扫描系统的变化，引入了又一个图像衰退的新来源。在成像过程中，对扫描系统有一个重大的挑战性要求，就是保持掩膜和晶圆台的同步移动，否则将导致图像模糊不清。通常的情况是一个方向上的扫描同步优于另一个方向，这将引起平行于扫描方向和垂直于扫描方向的线条宽度出现差异[9]。

掩膜版本身也是一个重要的工艺波动来源。制备掩膜采用的也是光刻工艺，覆盖在铬石英版上的光刻胶用电子束或者激光扫描工具曝光，显影后，曝光区域的铬膜被去除。掩膜版上发生的常见变化是平行于或垂直于掩膜写工具扫描方向上特征尺寸的差异、整个掩膜版上线宽的变化或平均特征尺寸的漂移。金属刻蚀工艺或者显影工艺的不均匀性均可能造成这些变化。

本节提及的工艺因素是工艺波动的常见来源，这些波动导致了生产制造过程中成像的偏差。其中一些波动是系统性的，在整个过程中变化不大；而另外一些工艺波动由于影响因素并不总是在工艺工程师们的控制之下，因而在整个过程中是变化的；但是有一点是相同的，即它们都会引起实际成像达不到可实现的最好成像性能。最终就对不同芯片之间电参数的一致性以及不同时期生产的芯片之间电参数的稳定性产生了影响。如果这些变化超过容限，将导致芯片失效。从芯片制造的角度来看，低 k 因子成像的主要影响是，为了减小工艺的波动程度，必须使用昂贵的控制机制。

3.2.3 低 k 成像对工艺敏感性的影响

在介绍了 90nm 节点对低 k 因子值的要求并讨论了工艺波动的来源后，下面分析低 k 成像是如何影响这些工艺的波动容限。我们将结合实例讨论剂量和焦距变化的容限以及掩膜版特征尺寸变化的影响，还要讨论 k 因子减小对图像逼真度的影响。选用的第一个实例是简单的一维结构，即光刻版上是节距相等的一组线条，对于不同的条宽尺寸，讨论光刻后线条关键特征尺寸的变化。第二个实例是二维问题，着重讨论低 k 成像对获得完美转角能力的影响。

3.2.4　低 k 成像和聚焦深度的影响

焦深（DOF）是指为了保持特征尺寸（CD）在可接受的 CD 容限内而允许的焦距变化范围。可采用下式对焦深进行简单估算：

$$\mathrm{DOF} \propto \frac{\lambda}{\mathrm{NA}^2}$$

上式也可以写作

$$\mathrm{DOF} \propto \frac{\mathrm{pitch}^2}{\lambda}$$

式中，pitch 为节距。上式表明，作为一个光刻参数，DOF 随工艺节点中最小尺寸的二次方变化。幸运的是，减小波长可以在一定程度上减缓这个情况。然而，如图 3-8 所示，随特征尺寸的减小，焦深的下降仍然很剧烈。对于 90nm 节点，焦深大致是 0.35μm。这应该与考虑所有误差源（其中一部分误差源前面已经讨论过了）所预期的焦距总变化估计值进行比较，参考文献［10］对此进行了详细的讨论。考虑所有误差源得到的总范围大约是 400nm，与基于前面方程计算的近似结果基本一致。随着工艺前进的步伐跨进 65nm 时代，期待的焦距变化将减小到 250nm。如果在不依赖于下一代 157nm 波长曝光的情况下能实现 65nm 的批量生产，减小焦距变化将是一个挑战。最重要的是，尽管晶圆尺寸一直在增长，减小由晶圆引起的波动的能力将是一个特别的挑战。

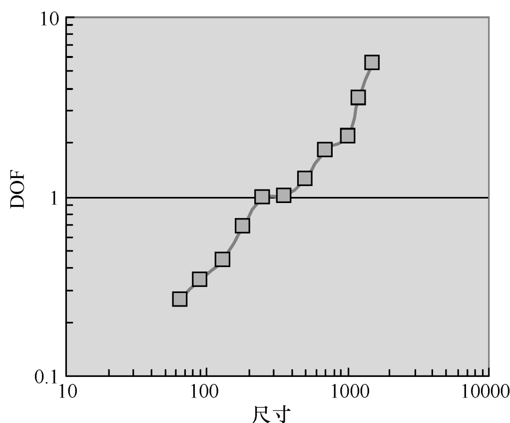


图 3-8　估计的焦深随技术节点的关系

3.2.5　低 k 成像和曝光容限

如上所述，低 k 成像通过两种主要机制导致空间像衰退：一方面，最终的像有一定的斜率，这是由于大于 1 的衍射级次对成像没有任何贡献。另一方面，由于对成像有贡献的衍射级次的部分越来越小，振幅调制将被减小。这意味着不管是由于光源本身变化或者是由于其下方衬底反射率的变化所引起的任何强度变化都会引起特征尺寸的变化。对于最简单的非相干光照明生成的线条状图形成像，可以根据像的斜率估计曝光容限范围。典型的光刻工艺应该根据要求的节距进行优化，并且调谐照明以优化对比度。曝光容限范围与不断减小的 k 因子之间的关系曲线如

图 3-9 所示，可以看出密集线条状图形的曝光容限范围随着 k 因子的减小而下降，与 193nm 工艺需要的曝光容限比较，大约在 10% 的水平[10]。

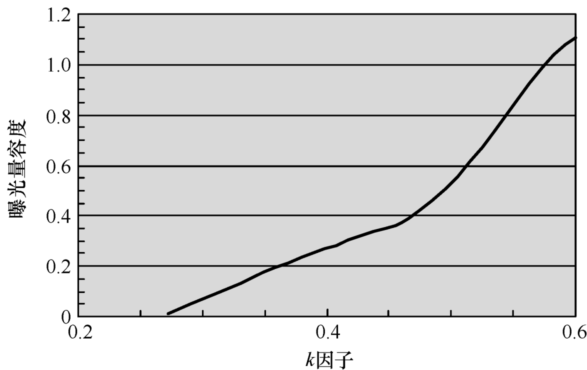


图 3-9　条宽与间距相等的线条图形情况曝光容限范围与 k 的关系
（曝光容限范围值是基于 0.75NA、193nm 常规照明曝光设备计算的，其中 σ 按照最优的曝光容限范围进行调整。计算的曝光容限范围仅基于空间像）

3.2.6　低 k 成像及其对掩膜误差增强因子的影响

掩膜误差增强因子（Mask Error Enhancement Factor，MEEF）描述了掩膜版上特征尺寸的误差与晶圆上实际特征尺寸最终误差的关系。MEEF 是晶圆 CD 误差与掩膜 CD 误差之比。MEEF 理想比值 1 之间发生的偏离是由于成像工艺的非线性造成的，并且将会随着 k 因子的下降变得更加显著。像许多其他的效应一样，MEEF 在高 k 因子成像时并不受到关注。图 3-10 定性地解释了 MEEF 值大

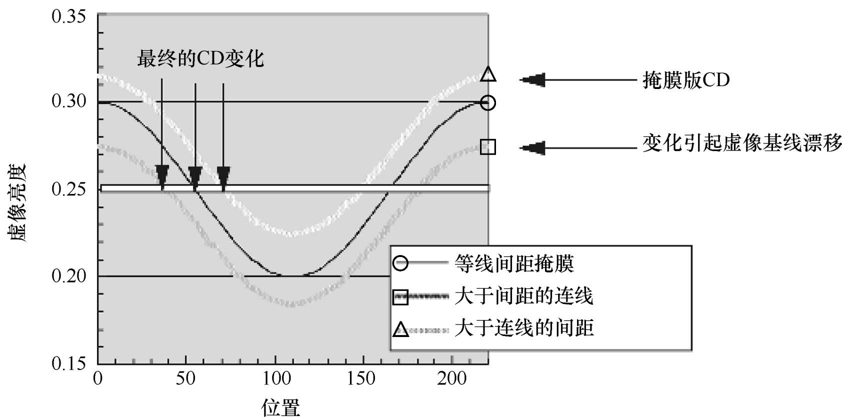


图 3-10　空间像和 MEEF
（掩膜版 CD 的变化引起基线的变化，进而转变为 CD 的变化。空间像斜率越小，CD 变化越大）

于 1 的原因。如图 3-10 所示，条宽与间距相等的线条图形的空间像是均匀背景和调制的叠加。我们主要关注条宽与间距相等情况的线条图形空间像的变化。假定掩膜版上的误差使铬膜线条宽度增加，在这一情况下，总的基线漂移来自于传输的平均光强漂移和调制的减小（第一衍射级强度减小；铬膜条宽和间距尺寸相等的情况下产生最好的传输）。空间像的斜率确定了交叉点位置基线漂移的敏感度。对低 k 成像引起的小的调制，即使较小的基线漂移也会导致关键的晶圆 CD 发生相当大的漂移。另一方面，较大的调制有更大的斜率，因此晶圆上 CD 的波动较小。该图也定性地表明，对于一定的节距，线条宽度与线条间距差别越大，产生的 MEEF 也越大。图 3-11 给出了条宽和间距尺寸相等情况下线条图形的定量估计。结果表明，采用 0.75NA、193nm 曝光工具，仅考虑空间像的贡献时，对非相干成像（$\sigma=0.9$），随着 k 因子的减小，MEEF 将增加。当 k 减小到低于 0.4 的范围时，MEEF 急剧增加。对于 CD 容限要求最严格的栅区光刻来说，这点特别重要，因为这时掩膜版 CD 误差对芯片线宽波动的影响最大。

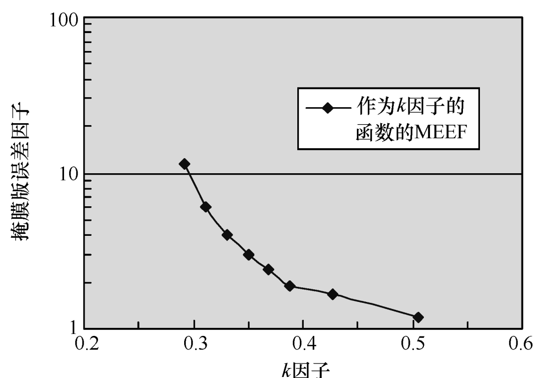


图 3-11　MEEF 与尺寸减小之间的关系

3.2.7　低 k 成像及其对像差的敏感性

以上我们一直把透镜作为理想状态来处理，在成像过程中未引入任何误差。然而任何使用过廉价放大镜的人都知道，用来成像的实际透镜决不完美。虽然光刻机采用的透镜的性能要远远优于只采用单透镜的放大镜（当然其价格也要随之贵好几个数量级），但是它们在成像性能上仍存在误差。这些误差被称作像差，可以作为使得衍射级增加了附加的相位差来处理，其大小与衍射角以及平面上衍射级次的方向有关。采用与波长之比定量描述像差的大小。现代透镜能实现 20 或更小量级的波误差。对 193nm 的曝光波长，相当于曝光工具只有约 4nm 的像差，这意味着必须制造和组装非常接近理想状态的透镜才能避免像差。当衍射角度更大时，波阵面误差一般会增加，这定性地说明当 k 下降时，像差敏感性增加。图 3-12 详细地说明了像差敏感性的增加情况。这些曲线是将像差类型和幅度作为随机变量通过蒙特卡罗仿真获得的。图中给出了仿真得到的特征尺寸并绘制了一个近似的包络函数。如图 3-12a ~ 图 3-12d，阴影区域不断增加，表明 CD

分散更加明显，这就证实了像差敏感性在不断增加。

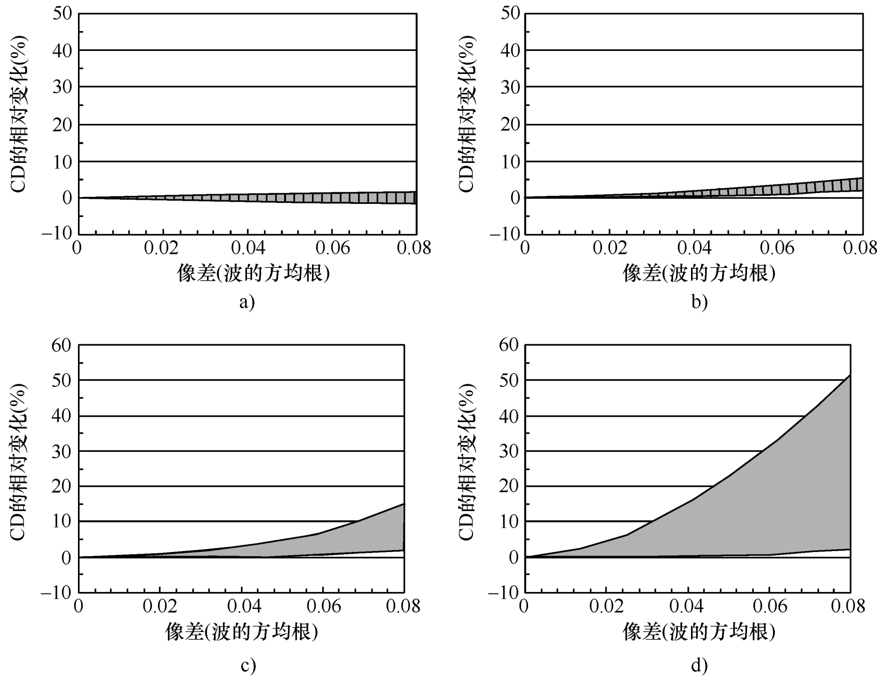


图 3-12 低 k 成像的像差敏感性

（采用与波长之比定量描述像差的大小。x 轴上显示的波前扭曲是由几种单独的像差组合形成的。对密集的线条图形情况，可以看到 CD 的变化范围与对波前扭曲有贡献的像差类型有关。当 k 因子减小时，该范围会激烈的拓宽，如图中四个实例所示。图中的 k 因子分别为0.6（见图 a），0.5（见图 b），0.4（见图 c），0.35（见图 d））

3.2.8 低 k 成像以及 CD 变化与条宽的关系

直到现在，仅讨论了由于工艺导致的波动。本节将介绍作为潜在波动源的物理版图的影响。本节的实例是一个简单的一维情况，即一个线条和间距规则排列的版图。其中只考虑间距发生变化，而线条宽度保持不变。图 3-13 重点说明了其关键效应：从较小的间距开始，由于低 k 情况下存在的非线性成像过程，使得光刻得到的线宽随着版图上线条间距的增加而减小。随着 k 因子的减小，该效应越来越显著。图 3-13 是基于 193nm 的曝光波长、0.75 的数值孔径和外部 σ 为 0.9 的常规照明情况下的仿真结果。每个曲线的最小节距对应于条宽和间距相等的线条图形情况，版图的半节距定义了 k 因子。每根曲线随节距变化的形状与数值孔径、所用照明的类型以及相干性有关。对于不同的照明设置，CD 与节距关系曲线的实例如图 3-14 所示。对于线条间距较小即线条比较靠近的情况，CD 的剧烈下降程度基本与照明的设置无关，这是由于彼此比较靠近的线条之

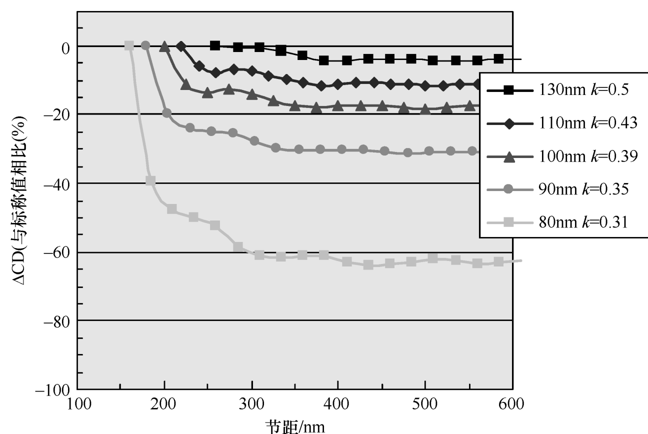


图 3-13　对简单的一维排列，CD 随节距的变化关系
（保持线条宽度不变，通过间距变化产生节距变化。对不同的 k 值，变化的幅度如图所示）

间的相消干涉引起的。而对于中等节距（大约 500nm 的节距）和长节距情况，CD 变化情况与照明设置的关系非常密切，这一效应有时被用来修正被包围的以及独立线条的偏移。尽管不同照明设置引起的效应不同，但它们都导致了光刻后 CD 的变化，而且变化程度与该线条距最近的相邻线条的距离长短有关。当试图减小光刻的最小间距时，该效应变得越来越显著。图 3-15 所示为最大 Δ 随 k 的变化关系。在低 k 范围，CD 变化已经接近线宽本身。为了对不同节距情况下能保持线宽不变，应在掩膜版上调整 CD 值以抵消这些变化。然而若修正较大，则易形成误差。用来修正这些成像变化的光学邻近修正技术将在 3.3 节中详细介绍。

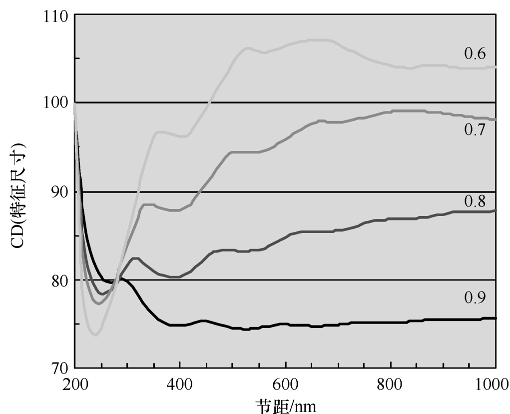


图 3-14　以照明设置参数 σ 为参变量，CD 随节距的变化情况

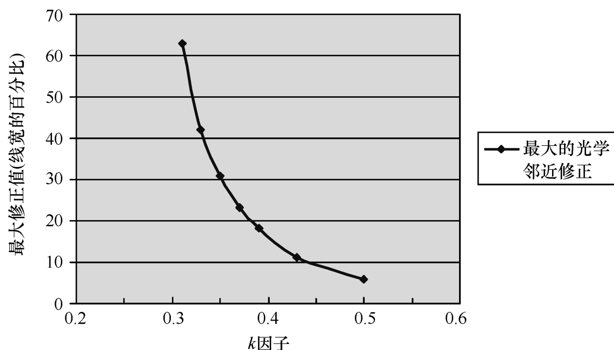


图 3-15　光刻后线条 CD 的最大偏差与 k 因子的关系

3.2.9 低 k 成像和拐角处的圆角半径

限制光学系统线性分辨率的效应也限制着光学系统对二维结构高逼真度成像的能力。由于成像的核心过程从数学角度描述，可为掩膜图像的傅里叶变换，这就解释了为什么空间像具有一些存在有限响应时间和信号衰减特点的信号处理电路的特性。对于输入方波信号，输出端信号拐角处变成弧形变化。当输入信号突变时，输出信号会出现减幅震荡。版图中拐角或转折相当于给电子电路提供了一个矩形输入信号。从较大块状图形中伸出较窄线条情况下的最终空间像的实例如图 3-16 所

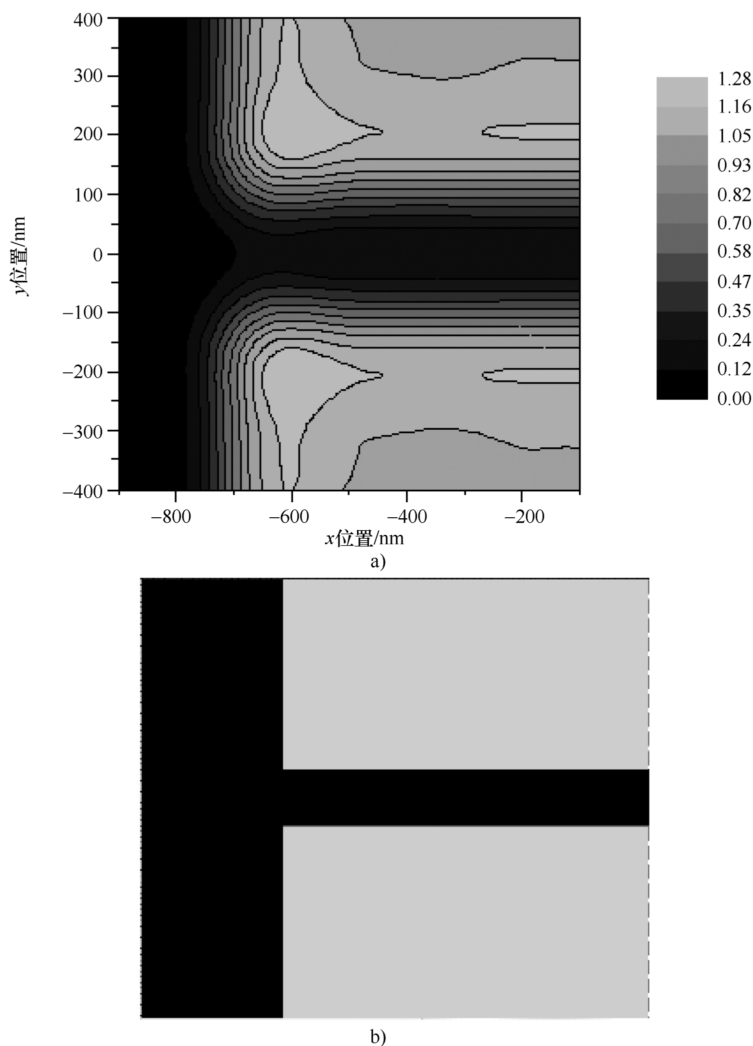


图 3-16 a）从压焊区这类较大块状图形伸出的较窄线条的空间像 b）相应的版图

示。空间像的轮廓并不像设计中那样有清晰的直角，而是在较窄线条从块状图形伸出位置的拐角处出现了圆角，离开拐角后线条开始变窄，但经过一段距离后线条却变得比标称线宽还要宽，最后在距离拐角较远的地方线条才恢复到固定值。图 3-17 定量地描述了这一现象，图中显示的是线条 CD 随不同位置 x 的变化关系。在 $x=-750$ 处，窄线条从铬版上较大块状图形中伸出。伸出后，在相当一段范围，空间像明显大于约 100nm 的目标 CD。直到线条伸出长度约 120nm 处，像 CD 才达到 100nm 的目标值，但是随之即下降到目标之下。线宽小于目标值的线条长度达到 200nm。接着线宽有小幅的增加。在其他一些特定设计中，例如位于两个栅之间的压焊区，也出现类似的现象。这些偏差对器件的影响很大。栅宽大于设计标称值使驱动电流衰减，另一方面，宽度变窄的部分则使泄漏电流增大。

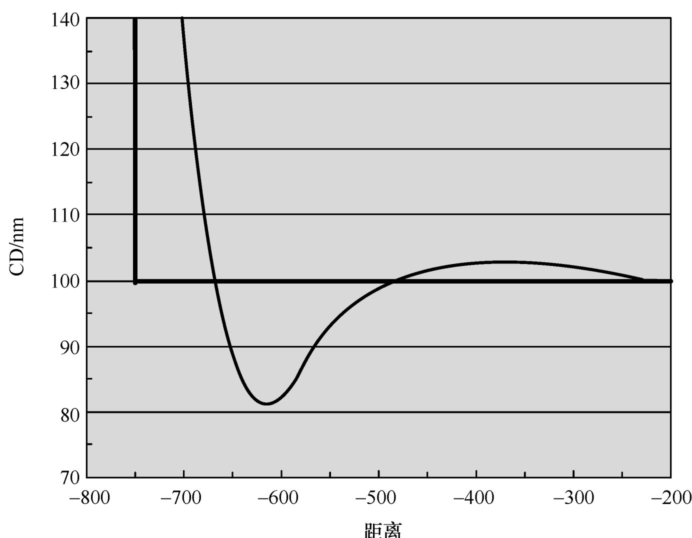


图 3-17　CD 与线条伸出距离 x 的关系（对应图 3-16 中的版图）

k 因子（或分辨率）对二维成像能力的影响如图 3-18 所示。图中比较了使用高 σ 值（$\sigma=0.9$）照明、193nm 曝光工具情况下，分别在 0.75NA 和 0.5NA 下形成的拐角空间像。高 σ 值照明将极大地减小拐角处出现的环形效应。0.75NA 系统的高分辨率能极大地减小圆角半径。低 k 成像对拐角处发生的倒角有一个附加的并且较为敏感的影响。例如，如果我们试着采用这种结构对节距为 240nm 的密集线间距图形成像，密集结构将确定成像的阈值。0.5NA 的成像阈值比 0.75NA 的低。然而，较低阈值（轮廓接近蓝光区）的圆角值比较高阈值的大，这将使性能进一步恶化。240nm 的节距结构对应 0.5NA、$k=0.31$，以及 0.75NA、$k=0.47$ 时的成像；两种情况下的圆角半径约分别为 210nm 和 130nm，即随着 k 的不同而变化。

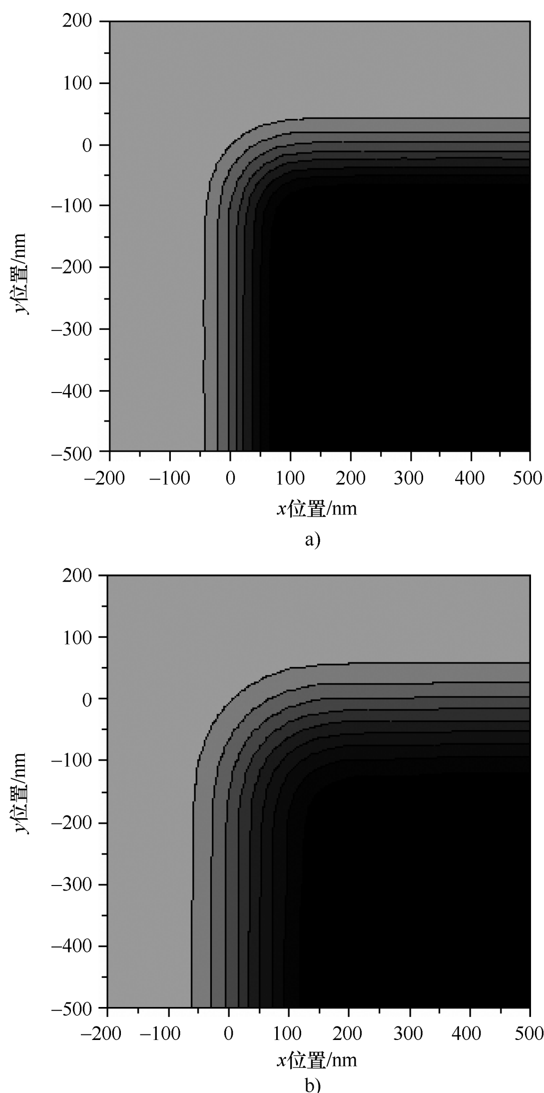


图 3-18 193nm、0.9σ 常规照明下的拐角成像

a) NA = 0.75 b) NA = 0.5

3.3 分辨率增强技术：物理

在 3.2 节，我们介绍了影响光刻工艺稳定性的几种效应。在低 k 成像时代，下述因素使得这些效应变得更重要：焦距深度范围的减小、低对比度空间像要求的曝光范围降低、像差敏感性的增加、以及一维和二维设计环境下图形逼真度的减小。为了应付这些效应，开发了多种技术[11]：(1) 专门的照明模式，(2) 亚分辨率辅助特性，(3) 相移掩膜版，(4) 光学邻近修正和 (5) 光学基本规则检查。

3.3.1　专门的照明模式

在 3.1 节曾指出，非相干照明的使用将相干照明情况下的分辨率极限（理论上的）从 $k=0.5$ 拓展到完全非相干照明情况下的 $k=0.25$。当接近极端低 k 的成像时，图像调制将减小，而较小的图像调制将引起各种各样的负效应。改善图像调制的任何方式都是我们热切期望的，并且像下面将要说明的，这也是使用专门照明设置的主要动机。采用衍射图形与照明光的卷积结果，可以很好地理解空间像调制的改善机制。通过透镜的衍射光决定了形成图形的调制，而照明光对衍射没有贡献，仅形成了一个均匀的背景，并使对比度衰减。图 3-19 所示为在大 σ 常规照明情况下密集间距成像最严重的缺点是相当的光仅仅起到增加均匀背景的作用，图形中仅仅交叉区域的光对调制有贡献，其余部分并没有携带任何空间信息。因此，去除照明光的中心圆部分是非常有利的，因为它去除了仅对均匀背景强度有贡献的大部分零级光，而或多或少地保持了对空间调制有贡献的部分。调整后的照明被称作环形照明，已成为大多数曝光工具的标准照明模式[12,13]。像常规照明一样，环形照明仍然是完全旋转对称的，如果版图对各个方向的图形均有要求，就要求照明具有这种完全旋转对称的特性。如果将照明的附加部分也去掉，就会得到更好的效果。一般的照明模式是四极或者偶极照明，分别对应于将整个环面分为四个（四极）或两个（偶极）部分。但是照明设备中这些部分的确切形状以及使用的名称互不相同，与设备生产商有关。简单的对称概念说明，在四重对称的四极照明情况下，仅 Manhattan 型几何形状图形才可以获得很好的光刻效果。与 0°和 90°方位图形相比，45°图形在成像特性上将会有最大的偏差。因此，为了利用经过改善的四极照明成像能力的优点，得到最小的基本规则，至少必须将版图限制为 Manhattan 型几何形状。如果 45°走向的线条尺寸较大，也是可以接受的。

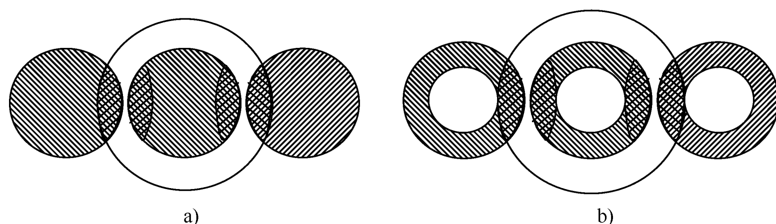


图 3-19　a）常规照明下密集间距线条图形的衍射图形（光瞳面中只有交叉阴影区域对形成图像有贡献。光瞳面内部的阴影区域只对统一背景有贡献。将早期的照明方式改为环形照明，可以减少对成像没有用的零级光的总数）b）使用环形照明后减少零级光的贡献

在讨论过的所有照明模式中，偶极照明的对称性最差，因此这种照明仅能对一个方向的线条成像就不足为怪了。对密集间距线条图形成像时，尽管目前在分

辨率极限情况下也还采用偶极照明方式，但是由于它仅能对一个方向上的图形成像，就严重地限制了其应用。人们曾尝试使用两块掩膜突破这些限制，其中一块掩膜是水平方向，另一块是垂直方向。遗憾的是，两次曝光极大地增加了光刻工具的工作量，因此就增大了制造成本。

定制照明模式具有其他一些优点，这将结合偶极案例加以说明。一个重要的特征是两束一级衍射光中只有一束穿过透镜，与传统照明中两个光束都有贡献的典型情况不同。这两种不同的情况被称作双光束和三光束干涉。双光束干涉足以完成空间调制。更重要的是，双光束干涉是唯一可以实现非常大的聚焦深度的一种方式。较大的聚焦深度只能发生在一级衍射光与照明灯的一极一致而零级衍射光与照明灯的另一极一致的情况。这种情况下，一级与零级光束有相同的入射角。遗憾的是，当图形间距增加时，零级和一级衍射光的入射角将不再相同，并且当间距偏离最佳间距范围时，焦距不易受影响的特性将削弱。

正如我们看到的，对非常特殊的版图情况，可以将特殊照明模式调谐到使成像性能最优的情况，而这将以牺牲一般图形的性能为代价。版图限制可能与方向和节距有关。因此，对于这种分辨率增强技术来说，版图选择将变成一个关键的使能因素。在有限的版图范围内，使光刻工艺具有一定的曝光和焦距容限。

3.3.2　光学邻近修正（OPC）

如 3.2 节中所述，低 k 成像导致光刻得到的图像与一维和二维版图的设计图像形状相比，发生了严重的偏差（见图 3-20）。主要效应包括：由于改变节距和线宽导致的一维图形线宽的变化、拐角处呈现的圆角和二维版图线条末端缩短现象。如果不修正，这些变化的影响是众所周知的。例如，栅极层次线宽的波动导致整个芯片范围线宽的变化（Across Chip Line Width Variation，ACLV）。在金属层次出现这种波动，将导致电阻增加。线条末端缩短将减小接触孔与下面金属互连末端的交叠，导致接触电阻增加。有源区内拐角处呈现的圆角将导致泄漏增加，或者使器件性能随交叠程度不同而发生漂移。

为了补偿这些效应，引入了一种叫光学邻近修正（Optical Proximity Corrections，OPC）的技术[14-16]。其名称很好地代表了这种技术早期的工作过程，即仅仅是基于两个互连之间的间距实现修正，同时也针对拐角处呈现的圆角进行修正。现今，这个名称使人产生误解，因为实际上除了专门的光学类型外，其他效应也能得到修正。

从概念上讲，OPC 是一种计算机算法，用它来调整原始设计掩膜上的图形形状。这些调整补偿了可能的成像失真，可以使最终在晶圆上光刻得到的图形尽可能地接近所期望的设计形状。图 3-21 所示为采用的修正类型作用示意图。通过修正掩膜版上图形的宽度对节距或者线条尺寸变化导致的线宽变化进行了修

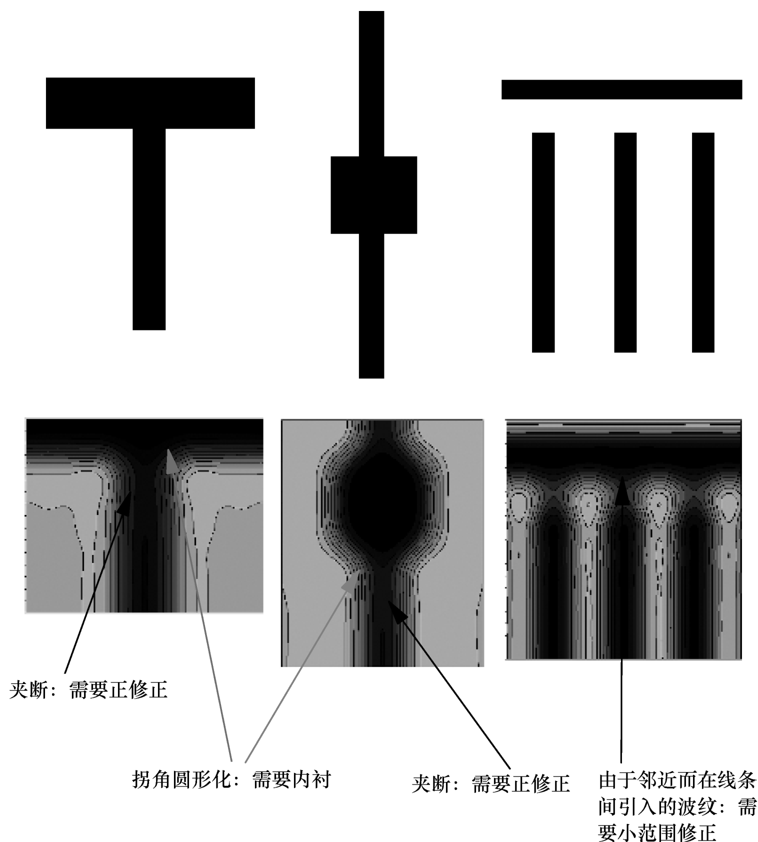


图 3-20　二维版图的实例及其 OPC 影响
（T 形连接线、接触压焊和梳状结构。图中位于上面是原始版图，下面的是最终的空间像，可以看见几个需要小范围 OPC 修正的变化）

正；通过在拐角外侧增加附加的图形或在拐角内侧去除部分图形可以对拐角处呈现的圆角进行补偿；只要在线条末端扩展一定量的图形形状就可以对线末端缩短进行修正，这本质上也等效于线宽的修正。从工艺的角度看，对修正圆角来说，将线条末端扩展和增加附加图形结合在一起使用效果更好。

晶圆上光刻得到的图形和原始设计的偏差不仅仅是由于光学效应，也包括光刻工艺中所有的操作步骤，从掩膜的制备、晶圆刻蚀和清洗工艺，都对图形失真有贡献[17]。对于栅极工艺，影响电学栅长的参数除了栅刻蚀工艺外，热处理、注入和侧墙工艺也对栅长产生影响。工艺中的这些问题将在后面讨论。

在某种意义上来说，掩膜制备过程与晶圆图形光刻过程相似，它们都包括了曝光和刻蚀步骤。例如，由于在掩膜版上生成图形的曝光设备分辨率有限，掩膜上也不能生成完美的拐角。这样，掩膜图形上就出现了会导致光刻线条末端缩短

的圆角现象。电子束掩膜写工具表现出了与光学工艺相似的邻近效应，这是由于高能电子撞击光刻胶和下面的铬膜造成的散射引起的。这导致了掩膜版上不同节距线条特征尺寸的变化，从而加剧了光学设备带来的光刻条宽随节距变化的效应。光刻胶也是影响临界效应、线条末端缩短和圆角的关键因素。对化学增强型光刻胶中发生的复杂化学反应很难进行建模。由扩散工艺造成的图像模糊仅是造成影响的效应之一。晶圆刻蚀工艺产生的效应与图形密切相关。例如，由于图形密

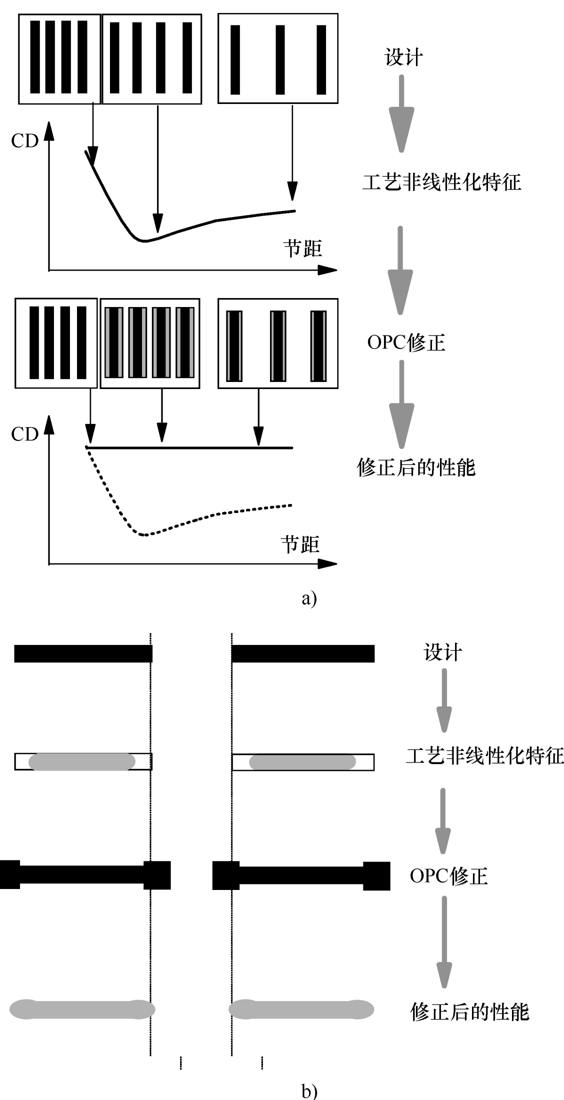


图 3-21　通常的图像失真机制及其 OPC 补偿

a）节距变化导致的成像 CD 变化　b）榔头状和线条末端缩短

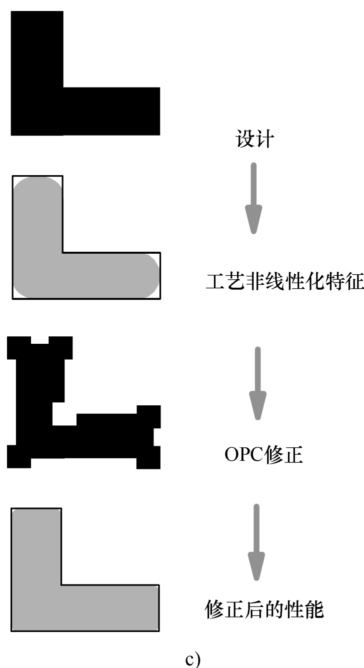


图 3-21　通常的图像失真机制及其 OPC 补偿（续）
c）内部和外部圆角

度、介质材料的局部带电和刻蚀过程中聚合体形成量的不同，均可能导致刻蚀速率发生局部变化。为了能够通过改变掩膜版上图形形状的修正算法对上述效应进行修正，这些效应必须具有系统性、可再现性和稳定性。随机的变化无法通过 OPC 来修正。

一般来说，任何一个关键工艺元素发生的变化，例如掩膜写设备的更换、晶圆刻蚀工艺的改动、光刻胶变化、甚至特征尺寸的改变，都将改变系统因素，并可能要求必须重新制备掩膜版。例如，在光刻工艺中为了获得更小尺寸的多晶硅栅而提高曝光剂量时，栅极线条 CD 随节距的变化以及 ACLV 均会加剧，这就在很大程度上抵消了期望的芯片速度的提高。

对给定的版图，可以很好地理解其光学成像过程，从而能够较快并精确地建模。而非光学效应则相当复杂，很难描述，通常采用现象学的方式处理。可以由版图得到一组便于通过计算进行管理的参数。基于这些参数，可以计算确定对空间像的修正。一般要采用实验数据建立这些参数与需要的修正量之间的联系。选择正确的、可以用最少的参数完成最精确的修正的一组参数是成功的关键因素，特别对那些容差要求严格的版图来说更重要。如果使用的参数太多，则收集必要的实验数据将变得很困难并且非常消耗时间。参数太少又可能遗漏了一部分重要的效应，从而不能实现正确的建模。

OPC 基本流程如图 3-22 所示。首先针对特定的版图，提取出一组参数，然后用 OPC 算法计算需要采用的修正，最后在掩膜版上形成一个经调整的新版图。需要的修正是基于之前收集和分析的实验数据而得到的。图 3-21 中还包括有版图的改进和晶圆图形的最终变化。一个 OPC 算法由几个关键的功能单元组成，其中一个单元的作用是将分层版图划分成多个模板。对每一个模板的修正都要单独进行计算，在计算过程中可以同时运行多处理器。然后就版图中的多边形组成部分进一步分成更小的部分，这个过程叫做分解。对如图 3-16 所示的二维效应进行修正需要这一分解过程。

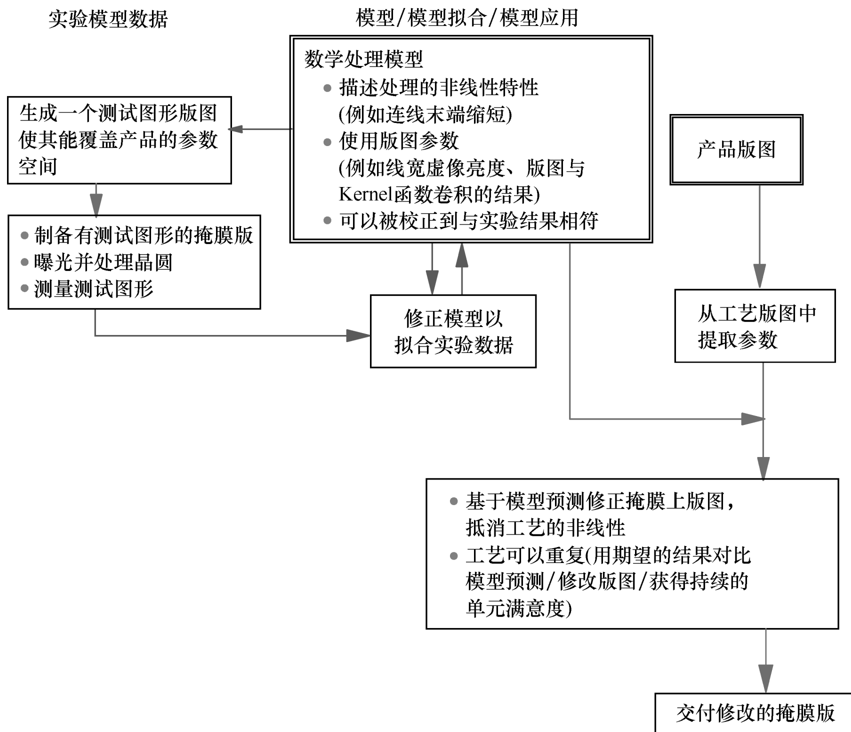


图 3-22　OPC 方法的简要流程图

OPC 流程的核心是用于确定每一部分移动量的算法。移动是基于期望的晶圆图像和它预计位置之间的对比，这个过程不断重复直到获得期望的精度。

可以采用两种不同的 OPC 方法，分别称作基于模型的 OPC（Model-Based OPC，MBOPC）和基于规则的 OPC（Rules-Based OPC，RBOPC），有时这两种方式联合使用。MBOPC 一般是基于光学图像计算以及对那些引起图形变形的非光学机制的物理分析实现修正，RBOPC 基本是纯经验的。例如，采用基于规则的方式时，首先收集如图 3-21 所示的对应不同节距的图形曲线实验数据。然后对

每种线宽与间距的组合，为了在晶圆上获得需要的宽度，确定需要的修正量。再在规则表中对应每个线宽和间距的组合填入修正量值。算法分析每个版图，决定线宽和相邻的间距宽度，再基于规则表中的数值确定修正项。遗憾的是，对于这种方式，即使最简单的一维情况，也会导致相当大而复杂的表格。线条末端和二维情况将会生成多组更复杂的规则。然而，我们应该认识到，对于简单的一维版图中修正线宽的变化而言，基于规则的 OPC 在任何的 OPC 方式中具有最高的灵活性，而且修正步骤明确、直观，适合于对修正进行多次精细的调谐。RBOPC 的主要缺点是规则设定巨大而复杂，并且设计规则使用的参数是基于设计尺寸和间距的，而这些对作为定义修正的参数而言并不十分有效。由于这些缺点，基于规则的 OPC 不可能在 100nm 节点及小于 100nm 以下使用。

基于模型的 OPC 的主要优点是它选择了比基于规则方式更有效的参数。对修正过程包含的物理理解用于建立计算修正的数值模型。这样，不仅使用了更有效的参数，而且也有定义明确的公式，覆盖了连续的参数值而不必将每个值分别列表。在参数空间选择的样本点采集实验数据，并在建模期间修正模型中的小部分变量。

基于模型的 OPC 方式最核心的是使用物理版图信息的卷积技术并且与一组预先确定的称为卷积核的函数进行卷积[18]。第 n 个核值 K_n 由下式给出：

$$K_n = \int k_n(x,y)\,\mathrm{layout}(x,y)\,\mathrm{d}x\mathrm{d}y$$

式中，k_n是第 n 个核函数。给出一组适当的核函数，卷积核方式最适合于计算空间像。除了可以用于简单的互连和间距图形外，更多地用于比较复杂的版图的建模。因为空间像是导致失真的关键贡献者，对于主要是由成像决定失真的情况下，这些类型的参数是一个很好的选择。因此，基于模型的修正在如图 3-16 所示的二维情况下表现很好，其中图像轮廓线从线条边缘处振荡展开。而这种情况采用基于规则的修正很难处理。可以说，处理这些版图的能力是针对较早的 OPC 方式进行的关键改进之一，并且对于像 SPAM 情况这类二维版图尤其重要。尽管处理非光学效应并不是设计模型时优先考虑的因素，而使用卷积核或者使用由空间像得到的参数都可以对很多类型非光学效应进行较好的建模。然而，建立非光学效应的合适模型对基于模型的 OPC 来说仍然是一个巨大的挑战[19]。

因为进行光学邻近修正需要进行大量的计算，因此不可能对设计中的每一点都进行调整，而是针对现有版图图形特点，在线条边缘处选择采样点。与采样点联系在一起的是具有一定长度（称为“链长”）的线段。基于对一点的计算结果，程序将移动一个线段的位置，在这段范围内采用相同的修正量。再基于互连间距的局部变化，添加一定的点数。另外，尖锐拐角处 CD 变化的实例也已表明，局部邻近变化已从实际发生变化的位置向外扩展了相当的距离。为了对发生变化的范围进行充分的采样，必须在发生突变或者变化的局部环境从变化点开始向外扩展增加大量的采样点。沿着版图上需修正的方向以固定的距离放置修正点

是最简单和最直接的策略。显然，线段的链长必须选择得足够小，才能适应局部环境可能出现的快速变化。然而，对于互相平行的延伸较长的线条没有必要采用同样的高采样密度，否则将导致过多的计算点和较长的运行时间。

针对局部环境的变化情况灵活确定并放置采样点是一种可行的替代方式。然而，这要求更灵活的搜索算法，不仅能检查图形内的设计变化而且能检查如前面指出的局部环境的变化。其中最具挑战性的问题是在线条末端合适地放置修正点以及相应的修正线段，这对线条末端的修正效果以及最后的光刻图形形状具有着重要的影响。设计中如何放置采样点以实现成功的形状修正与对工艺效应具有合适的建模能力一样重要。

基于模型的 OPC 是处理 3.2 节中分析的一维和二维图像失真所用的主要方式。不应该只是简单地将版图设计转化到掩膜版上，需要对转换到掩膜版上的图形进行有控制的且考虑周到的调整，其目的是消除失真。因此，进入低 k 时代后，基于模型的 OPC 已经成为一项关键的使能技术。它包含了对工艺效应进行建模的技术，使其成为优于陈旧的基于规则方式的关键优点。

3.3.3 亚分辨率辅助图形

亚分辨率辅助图形（Subresolution Assist Features，SRAF）是光刻工艺中扩展工艺窗口的常用途径之一[20-22]。不像光学邻近修正那样仅能对设计中存在的图形进行调整，亚分辨率辅助图形是加入了一些设计中不存在的图形形状。这些图形是基于对光刻工艺的详细了解，加入到掩膜中以提高掩膜性能的一种虚假图形，也称为辅助图形。通过分析环形照明对工艺窗口的影响，可以理解使工艺窗口得到扩展的关键机制。对于特定的节距，环形照明可以提供较大的聚焦深度增强效果。这个节距就是光瞳面中一级衍射光束和零级光束离中心的距离相同时的那个距离。尽管增强效果随着节距的变化而快速下降，但是在最优节距附近的一定范围内，还是具有较大的聚焦深度增强效果。为了将聚焦深度增强效果扩展到其他尺寸的节距，可以在原先的线条之间添加一些虚假图形，使最终的节距在选定的环形照明下更接近最优节距（见图 3-23）。在适当的节距中放置辅助图形，能给用于提高空间频率而采用的隔绝和半隔绝线条提供嵌套的环境，这就更好地符合前面提到的双光束条件。这个概念可以扩展到在每个间隔增加一个、两个和更多的辅助图形。文献中报道过曾经使用了多达四个辅助图形。从前面的讨论中也显然表明，辅助图形并不能用于任意小的节距中。只有当一级衍射图形在添加了新的图形后也能通过透镜，增加辅助图形才起到作用。应该注意，辅助图形一般并不产生新的衍射级次，但是可以增强已存在的高频衍射级次。由于只有在增强效应大到能足够克服极端分辨率限制情况下的调制损失时才表现出其优势，因此就制约了能使用辅助图形的最小节距。

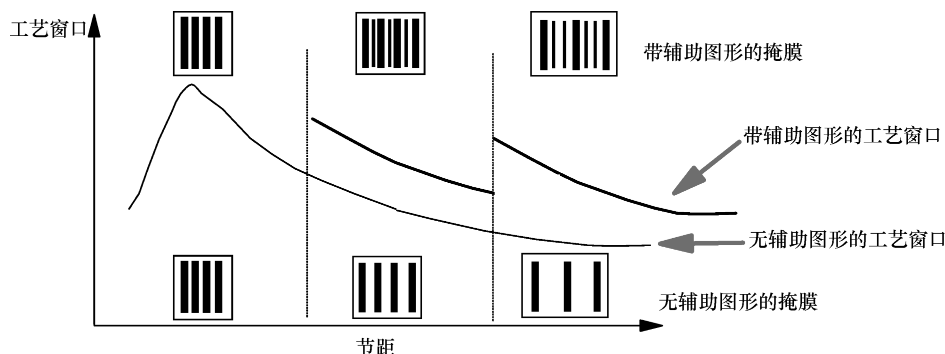


图 3-23 环形照明下工艺窗口与节距的关系，以及通过亚分辨率辅助图形实现的工艺窗口改善

如果辅助图形的宽度较大，它们将更加有效。然而，因为辅助图形在最终形成产品时又是不希望的，因此不应该出现在最终的光刻胶图像中。有两个可能的策略可以解决这个问题，其中之一是使用第二次曝光来去除辅助图形。然而不幸的是，这样的两次曝光将严重地减小曝光设备的生产效率、增加生产成本，因此一般是应该避免的。替代的方式是将虚假图形的尺寸减小到使它们不会出现在晶圆上。这是使用最普遍的辅助图形方式，同时也解读了“亚分辨率”名称术语的含义。但遗憾的是，基于可被光刻的考虑，其最大尺寸受到限制，我们还不能充分发挥其潜在优势。由于尺寸受到制约，因此衍射光强从一个级次转换到更高级次的效率受到了限制。辅助图形不会被光刻到晶圆上的最大尺寸取决于多种因素，首要的是 NA 和选择的照明类型、使用的光刻胶、以及版图线条和辅助图形之间允许的最小距离。避免辅助图形被光刻到晶圆上的要求也限制了在掩膜中给定位置可以插入多少辅助图形，可能必须执行避免插入三道和四道辅助图形的限制。基于掩膜的可制造性和检查的需要，也可能需要增加更多的限制。例如，可以规定辅助图形的最短长度。在布局之后，通过去除部分辅助图形使这些限制得到落实，但这就会导致一些图形或者部分版图没有被辅助图形保护。因此，设计的某些部分会出现比其他部分大得多的因工艺而导致的变化。这样，成功实施这项技术的关键就是使版图中尽量不要出现这种因为受到各种因素的制约而不能采用辅助图形的情况。

使用辅助图形的一个重要问题是与基于模型的 OPC 程序的相互影响。例如，按照辅助规则要求，将导致版图中影响 CD 变化的节距尺寸的不连续性。一般在放置有附加辅助图形的区域与没有放置辅助图形的区域之间，节距明显不同，光刻时需要的线宽修正量就会有较大差别。这种差别是由空间像的拓扑变化引起的。在未采用辅助图形的区域，沿着线宽方向，空间像的光强单调增强，在两条线条的中心位置光强最强。在添加有辅助图形的区域，由于在原来版图两条线条

的中心位置添加有辅助图形，使得空间像中该位置引入一个局部最小值。空间像分布的剧烈变化就可能给通过空间像参数来解决非光学效应的基于模型的 OPC 方法带来问题。因此，实施辅助图形就要求对模型进行非常认真的校准，以避免由 OPC 误差带来线宽的系统性变化问题。

3.3.4 交替式相移掩膜

交替式相移掩膜，有时也称作 Levenson 相移掩膜或硬相移掩膜，也许是最有效的一种分辨率增强技术[23,24]。毫不奇怪的是，无论是从掩膜/工艺的角度看，还是考虑到它们对物理版图的严重影响，它们也是实际使用中难度最大的分辨率增强技术。顾名思义，相移掩膜（Phase-Shift Mask，PSM）这个名称就预示着在该技术中利用了光波的相位特性。考虑图 3-24 所描述的情况，与穿过铬膜左侧的光线来比，假定铬膜右侧掩膜的透明部分引入了 180°的相移，则对位于这两个不同相位之间中间位置的图像，由于左侧和右侧光线传播的距离相同，但

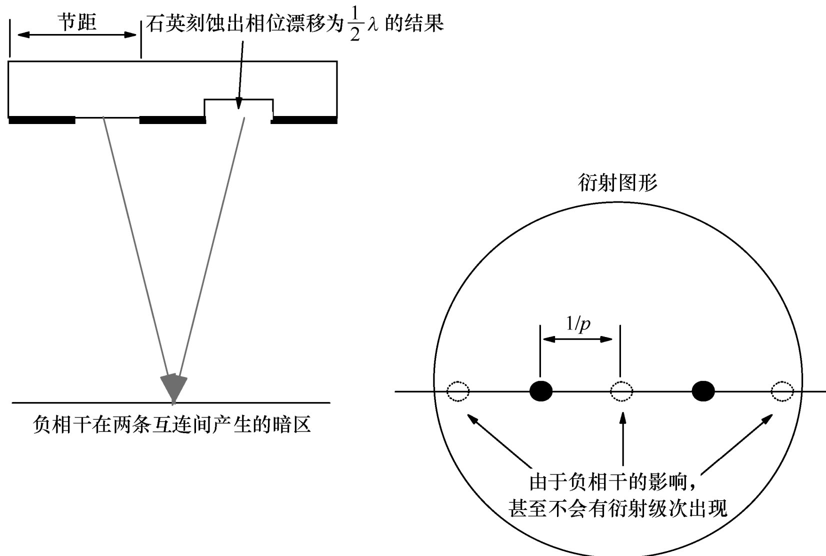


图 3-24 交替式相移掩膜的工作原理

（在交替式相移掩膜上的透光部分，部分区域石英呈凹陷状，另一部分区域没有凹陷。凹陷是对掩膜版石英衬底采用反应离子刻蚀方法形成的，一般接着还有一次湿法刻蚀。凹陷的刻蚀深度正好使穿透凹陷区域的光线相对于无凹陷区域有$\frac{1}{2}\lambda$的相移。这样，对于对称情况，在发生相移和未发生相移的透光区域之间发生了相消干涉，如图中下方所示，导致了铬膜上两个透光区域之间位置处的光强为零。应该注意的是，对暗区的产生，并不需要在两个相位区之间有一条覆盖铬膜的线条；在晶圆上相位不同的两个区域边界处将会形成光刻胶线条。对节距为 p 的设计，产生的衍射图形的间距为 $1/p$。但是，由于相消干涉，使偶数衍射级次被抵消了。更应该注意的是，不存在零级衍射。这样，不使用环形照明，交替式掩膜也可以实现双光束干涉的情况）

是它们彼此之间有 108°的相移，这两束光线的相消干涉就在两个透明区域图像的中心位置形成了一条全暗线。没有其他方式可以在两条间距如此小的线条之间形成暗区。在两个相位相反的区域之间产生暗线是一种干涉现象，即使去除掉铬膜上位于两个相位相反区域之间的线条，暗线还是会出现，这种情况称作相位边缘成像（phase edge imaging）。这种情况下，线条宽度完全由数值孔径、光强和选择的照明模式唯一决定。对交替式相移掩膜生成的衍射图形的分析，可以更进一步地了解工艺窗口的拓展机制。基于后面详细分析的原因，假定使用的照明模式是低 σ 常规照明。在使用简单的玻璃铬掩膜版时，产生的衍射级次之间的间距是(1/节距)的整数倍。然而，对于交替式 PSM，偶数衍射级次以及特别重要的零级衍射消失了。零级光束消失的原因很简单，因为中心图形左侧和右侧产生相移的透光线条宽度相等，而透过的光线相位相反，因此它们的光强相互抵消。这样，在仅有一级衍射穿过透镜的分辨率极限情况下，有两束干涉光，它们具有两个衍射级次并且距光轴的距离相等。前面曾经说明，在这种情况下，实现了较大的聚焦深度。与使用环形照明实现两束干涉光相比，交替式相移掩膜还具有其他优点。例如，提高衍射级次对称性的聚焦深度与节距无关，并且通过 +1 和 −1 衍射级次的干涉实现的最大调制远大于环形照明情况下的零级和其他级（如 +1 级次）的干涉。

交替式相移方式可以分成两类，即明场相移方式和暗场相移方式，如图3-25所示。如果采用暗场方式，则在设计线条的两侧放置宽度一定的两个移相线条。如果采用亮场方式，则在 0°的透光背景中放置一个 180°移相器线条。无论采用那种方法，采用移相版图均会新生成原来版图设计中所没有的图形。如果采用的是暗场方法，新生成的环形线条图形标示着 0° −180°的相位转换位置。如果采用的是亮场方法，将会在设计线条的四周新出现一大片有待移除的光刻胶。这两种情况下，需要额外增加一次曝光以最终得到要求的设计图形。这个掩膜被称作修剪掩膜或屏蔽掩膜，因为它修剪掉了不要的图形，而保留了要求的图形。

采用交替式相移方式需要对光刻工艺几乎所有的元素进行调整，要求新的掩膜制造、检查和修复能力。必须对曝光设备进行调整以有效适应双倍的曝光工作量。也需要能够针对已有的版图对要求的移相以及修剪掩膜图形进行综合编码，并能够考虑两次曝光编制光学邻近软件。最后，版图本身必须能够采用相移，这点将在 3.4 节分析。

使用交替式相移掩膜，需要进行两次曝光，将使曝光设备的生产效率严重下降。为了使这一问题减至最小，晶圆首先使用相位掩膜进行曝光，并在完成第一次曝光后晶圆仍然留在曝光台上。然后取下相位掩膜，换上修剪掩膜，转换曝光灯设置，采用第二块掩膜版对晶圆曝光后，将晶圆从曝光设备中卸载。为了减少掩膜版变换的次数，将下一片晶圆放到曝光台上后，先采用上一次晶圆曝光过程中的使用第二块掩膜进行曝光。即使如此，为了方便掩膜的更换，使延迟最小，

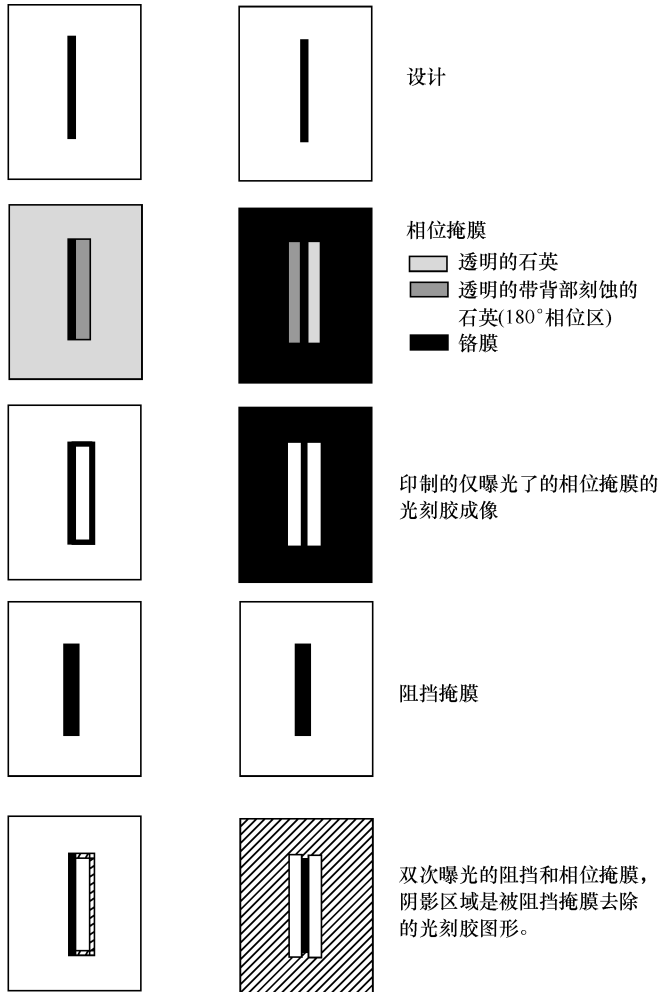


图 3-25　明场和暗场相移方式

(以一个简单的设计为例，图中显示了相移掩膜上的图形和使用正型光刻胶成像时光刻胶层上形式的相移掩膜图像。在明场掩膜方式下，应该注意，光刻胶层在相移线条图形的四周还新生成了环形线条图形，而光刻胶层上的这些环形线条图形是必须要去除的。同样，对于暗场相移掩膜方式，光刻胶层上新生成的较大区域图形也必须被去除。为此采用一个修剪（屏蔽）掩膜进行二次曝光以去除光刻胶层上这些不需要的图形。采用修剪掩膜可以去除不需要的光刻胶图形而保留了要求的图形。)

曝光设备的调整也是必须的。完成掩膜更换的机械部分必须能处理大量的更换而没有过多的磨损，掩膜版的对准和交换必须相当快，并且照明的变化也必须具有同样的速率。

比曝光设备调整更重要的是按照相移掩膜的要求改变数据编码方法。要求能够实现相位掩膜的编码方法。如后面将要指出的，除了可能的最简单版图外，只

能采用试探的方法。在期望的相位不能唯一确定的情况下，相位冲突是常见的问题。软件必须能综合必需的修剪掩膜图形。修剪掩膜一般同时包含有非关键设计图形以及与关键图形相关联的修剪形状。最后，OPC 算法需要具有仿真两次曝光的能力，主要是两次曝光光强的非相干叠加。因为涉及有两块掩膜，可以对两个掩膜或者其中的一个进行调整，因此必须有比现有的一般算法更复杂的算法。

最后，相移掩膜还要求调整掩膜的制造技术。掩膜的相移特性是通过刻蚀掩膜石英衬底中的部分石英决定的。在刻蚀石英之前，铬玻璃（COG）掩膜制造和交替式相移掩膜制造基本上是相同的：对覆盖光刻胶的无图形掩膜版进行曝光、去除需要被腐蚀掉铬膜的区域上的光刻胶、铬膜腐蚀、去胶、和掩膜清洗。在这之后，再次对交替式相移掩膜涂胶。紧跟着的曝光和显影步骤将去除掉需要进行一次附加的石英腐蚀以形成 180°相移区域中的光刻胶。完成 180°相移似乎只需要精确控制刻蚀深度，但是实际上实现这个要求并不是那么简单。考虑到透射过移相器位置石英窗口上的光线将形成衍射就可以体会到相移掩膜的光学复杂性。通过石英层的衍射光将被石英层下方的石英侧壁反射或被铬层阻挡。这个简单的情况表明，不仅对零级衍射而且对较大角度衍射的光线来说，实现 180°相移都是一个相当复杂的问题。其中石英侧壁的形状扮演着十分重要的角色。实际上，能够解决这个问题的仿真需要考虑掩膜的全三维结构，这也是需要重新进行广泛研究的课题[25]。

3.4　物理设计风格对 RET 和 OPC 复杂性的影响

随着光刻工艺向越来越小的 k 值推进，促进了对各种分辨率增强技术（Resolution-enhancement techniques，RET）的需求，包括特殊的照明条件、基于模型的 OPC、亚分辨率辅助图形和交替式相移掩膜。使用这些技术的能力越来越多地与选用的特定版图设计风格相互结合在一起。有两种方式可以通过 RET 来优化版图设计。其中一种方式是基于下述事实：对于给定的电路，可采用的版图并不是唯一的，而是各种方式都是可能的，并且经常是根据版图的方便程度选择相应的特殊风格。正是基于版图设计风格的多样性，可以使用 RET 优化特定的设计。虽然对设计者来说，不同的版图之间或许没有显著的优势，但是有时表面上似乎不明显的细节对能否使用 RET 却有决定性的影响。这些规则一般称为推荐规则、可制造性设计、规则或者类似的术语，通常被总结为导则而不是规则。如果不遵循这些规则要求不一定就导致灾难性的失效，可能只是使特定的版图对工艺波动缺少稳定性，因此规则所规定的要求本身涉及的边界经常并不是唯一的。

下面是版图/分辨率增强之间存在相互关系的第二个情况。特定的版图根本就与工艺团队计划采用的 RET 不兼容，或者采用 RET 将导致这些版图出现不能

容忍的严重的性能衰退。这些设计制约因素一般是设计团队和工艺团队之间的矛盾问题，这将构成数据预备流程的变化，下节将给出相关的实例。在以前的技术节点，RET 的选择问题基本与设计团队无关。版图设计团队提供了设计后即退到幕后，而由工艺团队确定并实施 RET，并且在这过程中对设计团队几乎没有反馈。在低 k 成像时期，这一问题不会再继续，RET 的选取对版图有显著的影响[26]。3.5 节将分析各种 RET 对版图的影响。大部分优化版图的基本原则可以通过一个简单的原理来加以理解：光刻能生成的最好图形是宽度和间距相等的一组线条图形。虽然在 DRAM 中出现的规模较大的线条阵列基本遵循这一原理，但是对随机逻辑，实际情况很难满足这一原理要求。

3.4.1 特定照明条件

低 k 成像的关键使能技术是特殊照明方式的选择。如我们所看见的，最有效的技术能减小照明模式的旋转对称性。按照在最可能的工艺窗口下实现最小可能节距的目标，对称性的减小是从照明中消除掉对图形形成没有贡献的光线的直接结果。这样，这些照明条件就有一个主要的缺点：仅对特殊几何形状的版图最有效。例如，采用四极照明时，与方位相对 Manhattan 类型几何图形成 45°的图形成像方式相比，Manhattan 类型几何图形的成像方式明显不同。基于四极的方向，将使图形中某一种取向的图形有较低的对比度，从而使 CD 容限以及工艺窗口的性能变差。假定四极照明已被优化为对 Manhattan 几何图形最有效，则 45°图形的工艺窗口和 CD 容限将明显变差。偶极照明的对称性更低，仅仅只有一个方向的线条能较好成像，这就对版图提出了十分严格的制约。然而，即使没有与偶极照明方式一起使用，采用单一方向线条版图最终生成的图形则具有额外的优点。由于不允许采用某一个栅方向的线条，就消除了导致栅线条宽度变化诸因素中的一个因素。如果两次曝光是可接受的选择，就无需只采用单一方向的版图，相应技术称作双倍偶极照明。遗憾的是两次曝光模式减小了曝光设备的生产效率，极大地增加了工艺费用，另外还增加了数据准备流程的复杂性。软件算法将图形中偶极方式的一个方向上的图形分开，形成掩膜版，单独光刻成像，而将其他的图形放在第二张掩膜上。在最简单的实现方案中，只是简单的基于它们的方向来分开图形。目前已经提出了更高级的处理方案，基于针对一对设计边界对比度的计算来划分图形[27]。对于每个设计图形，计算偶极方式两个方向的光学对比度，再按照最好对比度的要求放置在相应的掩膜上。

特定照明条件的使用也意味着对于一定节距的图形其成像效果最好。图3-26所示为由线条和间距组成的简单结构图形阵列的工艺窗口，其中互连宽度固定，但是间距变化。如果使用环形照明，图中画出了工艺窗口与间距之间的关系曲线。在线宽等于间距这一条件附近一个相对较窄的区域，存在优化的工艺窗口，

随着线宽与间距的差别越大，工艺窗口急剧退化。在正常工作条件下，较小的工艺窗口一般会导致较差的 CD 控制，甚至产生更差的、灾难性的失效。另外，这些区域一般对应较大的 OPC 修正区域。其结果是使用这些节距的版图将增大对 OPC 建模过程中误差的敏感性，线条宽度控制效果也较差。解决这些问题的一个可能的有效途径是引入禁止区，即对比可以获得的工艺窗口与最小需要的工艺窗口来确定禁止区。最小需要的工艺窗口来源于工艺团队对工艺波动的详细分析。基于最小需要窗口和图 3-26 这类图形，就可以确定那些不满足要求的节距。禁止区节距的细节、范围和位置与使用的光刻方式类型密切相关，并且在各个加工厂之间也不会相同。对 0. 13μm 和 0. 1μm 技术节点的详细讨论可参见参考文献［29］。从图 3-26 那样的图示曲线中还可以看到的另外一个结果是，如果一条单独的最细线条紧挨着很宽的间距，则存在工艺窗口严重化的问题。如果是一条单独的最细间距紧挨着很宽的线条，情况相同。一条较宽的电源总线与一条其宽度与电源总线相同的较宽地线之间采用的是最细间距相隔开的情况，对光刻和化学机械抛光来说是一个众人皆知的存在问题的领域。光刻团队通常没有选择，为了维持产量只能放弃他们的 RET 选择，因此更希望从设计角度采取解决方案。大部分情况下，即使略微增加最小宽度（10% ~2%），也将得到明显的改善。

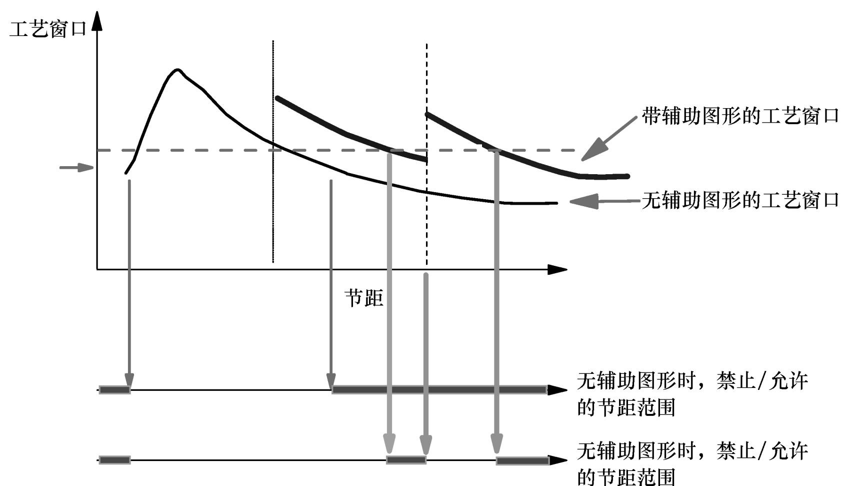


图 3-26　由线条-间距组成的图形的工艺窗口，其中固定线宽，节距变化
［图中同时显示有代表最小工艺窗口要求的线条。工艺团队可以基于这一要求建立禁止区
（即在版图中为了实现最优性能而应该避免采用的节距）］

亚分辨率辅助图形经常与环形照明联合使用，可以在很大程度上改善工艺窗口性能，并对所有的间距可以促使工艺窗口维持在最小要求之上。即使这样，对非常低的 k 值成像来说，工艺窗口问题仍然会存在。对于辅助图形工艺，在节距轴上会多处出现禁止节距区域。这是在节距轴上辅助图形的数量以十分规律的间

隔增加的结果。对于节距略小于能引入额外辅助图形的节距的情况，工艺窗口最小。即使在整个节距范围，工艺窗口都足够大，如果试图使 CD 波动足够小，由额外增加的辅助图形所引入的空间像不连续问题对任何 OPC 模型来说都是一个极大的挑战。因此，对于像栅极这样的关键层次，有一种趋势是从允许的节距中排除这些转换区。然而，节距限制不再是单个层次的问题。一组完整的基本规则必须能适合于提供一个有用的版图框架。图 3-27 所示为多种基本规则的调整实例。在这一常见的版图实例中，两个栅被放置紧挨着有源区的一个拐角，接触孔放在两个栅之间。在这种安排中，为了避免像图 3-27 所示的版图那样导致所需版图面积的过多增加，许多基本规则，如栅的最小宽度、最小接触孔尺寸和图形的相对放置必须与允许的栅节距一致。迄今为止，我们的考虑基本上是针对一维版图，并且讨论了与工艺窗口有关的问题和 OPC 能力。从设计的观点来看，在面积方面能做的最重要的改善是尽可能地维持设计图形的单一方向，并减小不同的节距值数量。

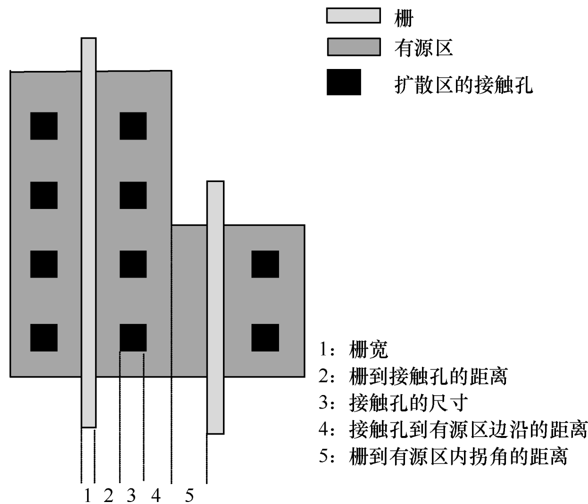


图 3-27 栅节距、接触孔尺寸、最小栅接触孔、以及接触孔与有源区边缘最小间距等设计规则之间相互关系的版图设计实例

（这些基本规则必须与 PC 禁止节距规则相匹配，以防止过大的版图面积）

3.4.2 二维版图

对低 k 成像来说，一组有关的附加问题是需要考虑二维版图。在许多情况下，它们损害了光学邻近修正的能力，并且一般会减小几乎所有 RET 的效率。如我们所看到的，与低 k 成像有关的几个效应导致了下述结果：线端缩短、拐角变圆、以及线条宽度突然较大变化引起图像的非局部减幅振荡等对短而直的线条提出了极大的挑战，而对长的直线条，这些效应的关注很少。基于模型的 OPC

已经极大地改善了这方面的能力，虽然修正局部的线宽变化仍然是一个挑战。图 3-20 显示了版图二维问题对 OPC 很重要的各种版图类型。一个实例是由两根线条成的 T 形连接，特别是在 SRAM 中十分常见的“腿”很短的 T 形连接图形情况。另一个实例是在两个较窄的器件之间放置的接触压焊点情况。最后的例子是一系列的水平栅附件存在一个垂直栅。通过分析一个未修正结构的空间像，就可以了解与这些设计相关的问题。T 形连接图形显示出先前讨论的空间像的波动性。因此，在离开 T 形连接处的位置存在线条被夹断的趋势。在这些区域，OPC 算法通过加宽线条试图修正线宽的变化。OPC 残存的误差和不完美的片断会导致对线条上相应部分线宽的控制变差。如果该线条是有源栅的一部分，过度修正将导致驱动电流减小，而修正不足将导致泄漏增加。如果其中一段线条较短，形势将更加严峻。这种情况下，OPC 修正不仅必须补偿空间像的摆动，而且必须补偿较短线条段上的线端缩短。在短线条段端头的榔头形状更一步增加了空间像的摆动。为了以足够的精度修正这些效应，所有的结果将是高度破碎的掩膜像，这不仅增加了数据量而且本身也易于产生误差。因为器件相当窄，由泄漏或驱动电流表示的性能将十分易变。与此相似的一种情况是在两个较窄器件之间的接触压焊点。器件中心部分宽度的变化导致有源栅区发生摆动。如果栅很短，端头的榔头形状将在其中一个端头上由接触压焊点引起对空间像的另一个干扰。如果使线条末端和接触压焊点保持在发生空间像摆动的区域之外，就可以改善线宽控制。

下面通过几个实例说明使用短线条时 RET 效率的减小。图 3-28 所示为版图实例，该设计包括了间距为 660nm 的一组线条，采用 193nm 曝光和 0.7NA 的环形照明成像，采用亚分辨率辅助图形来提高这些半隔离线条的工艺窗口。在相当长的线条之间放置了两个辅助图形，每根线条的目标 CD 为 90nm。图 3-29 所示为图 3-28 中版图的工艺窗口。对相当长的线条，通过使用亚分辨率辅助图形使得聚焦深度增加接近 70%，如图 3-29 中虚线所示。图 3-29 中的实心方块对应

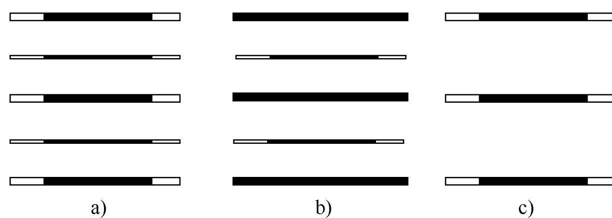


图 3-28　短线条情况下亚分辨率辅助图形效率降低的实例

（这些版图被用于图 3-29 中的计算）

a）采用辅助图形的版图：辅助图形与主线条的长度相同

b）采用辅助图形的版图：仅将辅助图形的长度缩短

c）未采用辅助图形的版图

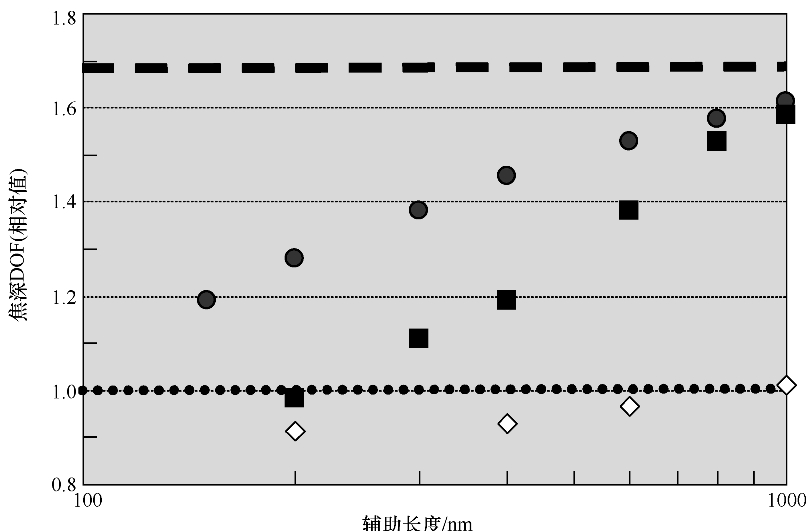


图 3-29 对图 3-28 中的版图，随线条长度变化而变化的工艺窗口减小情况（本图说明短线条情况下使用辅助图形的优点在减小，主要的效应是辅助图形变短则益处减小，实心方块对应图 3-28a；圆对应图 3-28b；菱形对应图 3-28c）

辅助图形和版图线条逐步缩短的情况。工艺窗口优势在特征长度为 1μm 时开始下降，并且在线条长度为 200nm 时完全消失。即使在 400nm 时，线条长度可以获得小于总工艺窗口增加量的 1/3。关于问题的进一步研究表明，DOF 损失是两个效应的综合结果，其中一个效应是当辅助图形变得更短时，它的有效性将减弱。图中实心圆表示的是中心线条的长度保持相当长，只是缩短辅助图形的长度，完全证明了辅助图形缩短的效应。第二个因素是主要图形的长度，在图3-29 中用空心菱形表示。在这种情况下，没有使用辅助图形，仅仅使线条长度缩短。这时工艺窗口损失适中但仍然存在。这个实例说明，短线条图形对包括亚分辨率辅助图形或专门照明条件这类 RET 是不利的。

应该注意，通过采用辅助图形获得的工艺窗口增大以及随着线条长度日益减小而发生的工艺窗口衰退的严峻程度和速度的细节情况，与多种因素有关，包括节距、照明条件、辅助图形、和辅助图形的放置等，并且变化很剧烈。然而，短线条图形对有效的 RET 是不利的这一基本结论仍保持未变。

上述作用基本与这样的事实有联系：当线条长度减小时，由线条阵列确定的衍射级次的局部化正变得日益减少。另一方面，RET 优化衍射级次图形的工艺窗口性能则在光瞳面的局部区域。当图形中越来越大的部分位于亚分辨率增强技术的优化性能区以外时，它将变得越来越无效。大多数 RET 的工作是基于由线宽与间距相等的线条组成的阵列，具有固定节距的长线条导致高度局部化的衍射

级次，这是 RET 工作效果最佳的情况。版图越接近这种情况，就对 RET 越有利，OPC 修正的应变也就越小。如果版图设计中包含有许多弯曲和扭曲的短线条图形，意味着版图将导致 OPC 程序发生巨大的应变，采用 RET 几乎不会得到什么益处。

在多数情况下会产生短线条图形，而且不同情况下的结果差别较大。图3-30 所示的版图实例将导致采用短条辅助图形出现的问题。当水平线条阵列被辅助图形保护得很好时，对垂直互连的保护则很差。辅助图形生成程序将尝试按照图中所示的情况放置辅助图形[30]。水平辅助必须被削减以避免接触主要图形。由于下述事实，使得问题进一步复杂化：放置辅助图形时，通常不允许出现三路交叉，否则交叉将可能不再成为“亚分辨率”，导致晶圆上残存光刻胶。如果中心的水平互连没有辅助保护，它将变成一个严峻的工艺问题。如果像图 3-30 中那样的水平线条不能避免，将垂直线条末端和水平线条之间的距离增加到至少可以放置一个辅助图形当然是一个选择。然而，这却导致了面积明显增加。在这种情况下，最好的有效利用面积的方案通常是把中心水平互连的宽度改变为非临界互连宽度，从而无需采用辅助图形来保持要求的工艺窗口。

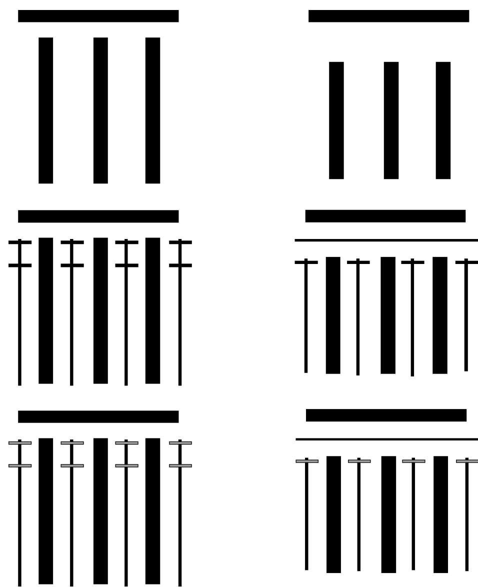


图 3-30　两个梳状结构的辅助图形放置方案比较

（第二个例子中的水平线条与垂直线条之间的间距已经增加。最上面一排是原始版图；中间一排是辅助图形初始放置情况；最下面一排是在采用清除码之后保留的辅助图形。左边的实例中，清除码必须去除所有的垂直辅助图形，因为它们会产生不希望的辅助交叉。在右边一列中，间距已增大到可以放置一个连续的水平辅助图形。左边一列导致严重的工艺困难。水平互连不仅没有辅助图形，从而导致极差的工艺窗口，而且由于强烈变化的近邻，使 OPC 处理复杂化）

3.4.3 交替式相移掩膜

交替式相移掩膜对物理版图有非常大的影响。可以使用交替式 PSM 的图形与不能使用交替式 PSM 的图形相比，工艺窗口改善是所有 RET 中最显著的。如果版图设计采用的是最小尺寸，绘制时必须采用相当大的倍数，但是却不能进行相移，则这种版图一般是严重影响成品率的图形。由于使用交替式相移掩膜增加了不能被相移的版图的“似然性”所带来的一系列新问题，使得这种非相移版图的严重问题进一步恶化。为了理解这些复杂性，下面首先分析与交替式相移掩膜相关的基本规则[31]。

每个关键线条与两个相移区相邻，零相移区在线条一侧，180°相移区在另一侧。各种规则通常用于决定每个相移区的最小和最大宽度、以及两个不同相移图形之间的最小距离。仅在相位相同的区域之间允许交叠，因此只有在采用并列交叠排列的情况下才可能使放置密度最大。最简单的规则是考虑到掩膜制造工艺要求符号相反的相位之间距离最小。从工艺角度考虑，将决定允许的最大相位宽度以及不需要采用相移的最小线宽。图 3-31 说明了这些规则的基本设计含义。此外，还有一组规则用于防止在设计中使用了导致相位冲突的几何图形。相位分配是在设计中按照线性路径进行的。关键线条一侧的相位意味着该线条另一侧的相位相反。相位冲突是因为针对特殊间距安排的相位相互矛盾，这一情况与选择的是哪条路径到达相位区有关。导致相位冲突的两个常见版图实例如图 3-32 所示：关键 T 形连接以及奇偶路径情况（也称为“皮带扣”式版图）的一种图形。这

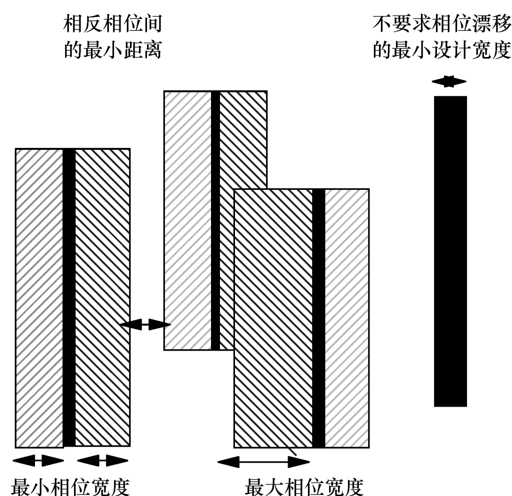


图 3-31 交替式相移掩膜的基本版图情况（图中显示了交替式相移掩膜版图设计的基本规则，包括最小和最大移相区宽度、相反相移图形之间的最小距离及无需采用相移的最小线宽）

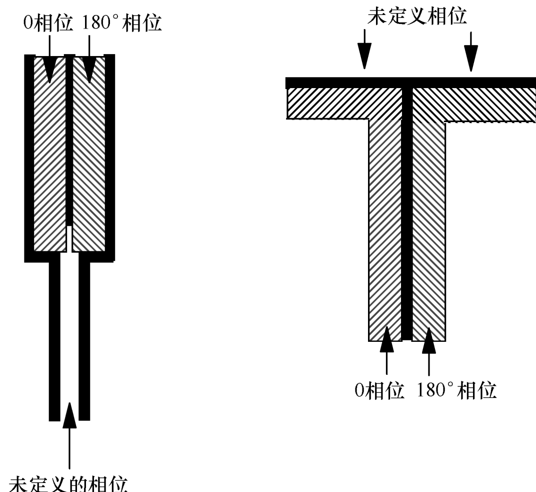


图 3-32 导致相移冲突的版图情况：奇偶路径和 T 形连接

些版图仅仅是常见的设计误差的简单实例，又称为（圆）周期误差，这些误差与相移掩膜有关。图 3-33 所示为另一个实例。相位分配按顺序进行。在图 3-33 所示的版图实例中，在版图的不同位置，这些路径分开后又重新组合在一起。在路径分开的位置以及重新组合的位置之间，每条路径已经跨过的线条数目可能是偶数或奇数。如果一条路径上有偶数个相位转换，而另一条路径上有奇数个相位转换，则在路径重新组合的位置就不可能实现唯一的相位安排。这组失效着重说明了基本规则布线检查面临的重大挑战。由于误差能在相邻的图形之间被检查出，误差就不再是局部的了。

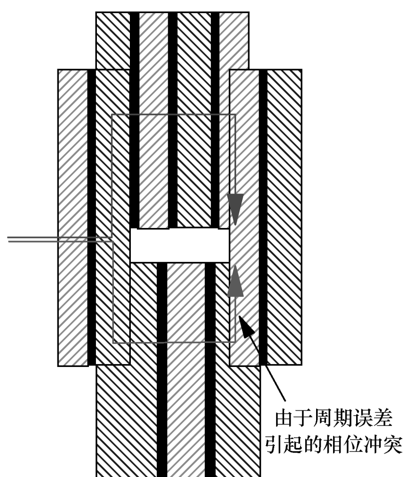


图 3-33 另一种错误的实例
（随着路线分岔并且最后又结合在一起，这些错误就产生了。上面一条路径穿过三根设计线条，而下面一条仅穿过两根。这些错误对设计规则检查提出了巨大的挑战，因为它们是非局部的）

与解决相位冲突相关的困难导致了设计和光刻工艺之间相互作用的典型转变[26]。先前的技术在设计空间和工艺空间之间有一个清晰的分隔，并且大部分情况下采用的 RET 对设计并没有影响。使用 RET 引起的限制和工艺关注表现为一组设计规则限制，与前面章节讨论的情况相似。相位冲突问题的复杂性驱使方法学上的变化。一种方案是把相移工具转移到设计

团队手中，使他们具有检查相位分配的能力，甚至还可能检查预期的掩膜修剪形状。这样，确保交替式相移掩膜灵活性的责任就转移到设计团队手中。考虑到有时相当复杂的交替式相移规则框架，再结合标准的基本设计规则，误差的识别可能成为令人灰心的经历。需要指出的是，交替式相移规则不仅仅包括相位分配规则，还包括一些附加规则。例如，描述相反符号的相位形态之间的最小距离、相位形状交叠的规则、以及最小的修剪掩膜尺寸等等。因为这些参数可能非常特殊，它们不仅阻止了在不同制造厂之间的交流，而且也不能用于更先进的光刻方法。对 ASIC 设计方式来说，在两个层次上要求对相位的适应性。第一，单个库单元的版图必须具有相位适应性。一旦这个条件满足，当组合使用单个库单元时，还将出现更复杂的问题。为了避免这个相位误差，必须通过在单元边界周边提供足够大的缓冲区以保持单个单元相位不受外界影响。通过尝试寻求相对简单的单元边界特点，可以采用相对简单的并列交叠步骤将 ASIC 设计中成行排列的单元放在一起，就可以减缓这一需求。参考文献［33］对这种方式进行了讨论。

另一种替代方式是建立设计制约规则，可使得设计团队无需考虑相位分配的复杂性以及修剪掩膜版图，但是对传统的设计过程来说冲击则是明显的[26]。在这个方式中，有时也被称作粗网格方式，设计被局限在一个方向，并且接触节距只能采用整数倍的变化。这种框架避免了复杂的二维版图情况。这种方式有几个极具诱惑的优势。它对多种 RET 提供支持，解决了由二维情况引起的可制造性问题，并且当转移到新的技术节点时，能减少重新设计的需要。这种方式的主要挑战是保持面积紧缩到最小，并且为了实现栅极层次版图的简单化，应该尽量减小导致其他设计层次版图（例如局部互连层次）复杂性的增长。

如前所述，在光刻工序中更多地考虑 k 因子就会改变设计和数据过程之间的关系。在先前的技术节点，主要考虑生成最小尺寸的设计图形，包括互连、节距或面积，以及决定不同层次关系的一组规则，使覆盖容差不致导致芯片失效。由于 100nm 和 70nm 节点技术更加依赖 RET，为了充分发挥新技术的作用，需要引入一组新的设计规则。图形方向的限制、从设计中清除特定间距范围的禁止区、以及线条图形长度的最小限制等方面的要求将日益严峻。交替式 PSM 将对物理版图有一个更加重要的影响。特别是栅极层次，由于对 CD 容限有严格要求，在这个领域处于领先位置，而其他层次也跟随其后，特别是在 70nm 节点。由于 Foundry（代工厂）生产线存在巨大的竞争性压力，生产方将极不愿意把这些限制加到他们的客户身上。设计团队应该非常清醒地认识到，光学成像物理、曝光设备和 RET 的实行在业界已经相当标准化了。一个了解相关方面知识的客户（即对低 k 成像实施有清醒认识并且明确如何优化版图以充分发挥光刻 RET 效率的团队）将通过使设计更加适应工艺的波动而受益，因此更可能提高成品率。对设计团队来说，益处是显而易见的，光刻性能的重大改善将最有可能缩短产品上市时间。

3.4.4　掩膜版成本

早期的光刻，掩膜版的开销在芯片制造总成本中占的比例并不大。但是现在这一情况已经发生了巨大的变化；对 90nm 节点的产品，一组掩膜版的费用在一百万美元的级别[34]。掩膜费用是光刻工艺成本中增长最快的部分，超过了晶圆工序和曝光设备费用的百分比增长。对 ASIC 设计，由于只要加工为数不多的晶圆就能满足客户的需求，使得这一问题更加突出。由 Sematech[36] 对其成员公司完成的调查表明，一组掩膜版曝光的平均晶圆数，对逻辑 ASIC 来说大约是 500 片，而对 DRAM 生产，大约是 8000 片。基于业已报道的大量 90nm 技术节点 ASIC 设计数据，经过外推处理，结果表明，掩膜版开销已占到光刻总开销的 60%[37]。

对低 k 成像的需求以及为此采用的 RET 是促使掩膜版成本增长的主要因素。驱使掩膜版成本增加的主要因素有两个：掩膜版制备过程中的写时间和掩膜成品率。这里所说的掩膜版成品率是包括最小缺陷标准和 CD 要求在内所有规范要求的掩膜版成品率。图 3-34 总结了由于使用低 k 成像和 RET 使得掩膜版成本增加的因素和机制。从底部开始，列出了向低 k 成像时代推进的各种因素。其中一部分因素是任何工艺缩放过程所通有的。每个掩膜版范围内图形数量的增加和更加密集的成像布局要求对各种基本规则的减小都是存在的，在此将其列出只是考虑到完整性。这两个因素都影响掩膜写时间。由更密集布局提出的精度要求，增加了载物台的移动次数，因此增加了写时间。增多的图形数量也增加了需要被写入的数据总量。因为先进的掩膜写工具是半导体工业加工设备中最昂贵的部分，而增长的掩膜写时间将增加掩膜开销。如 3.2 节所指出的，低 k 成像增大了掩膜误差

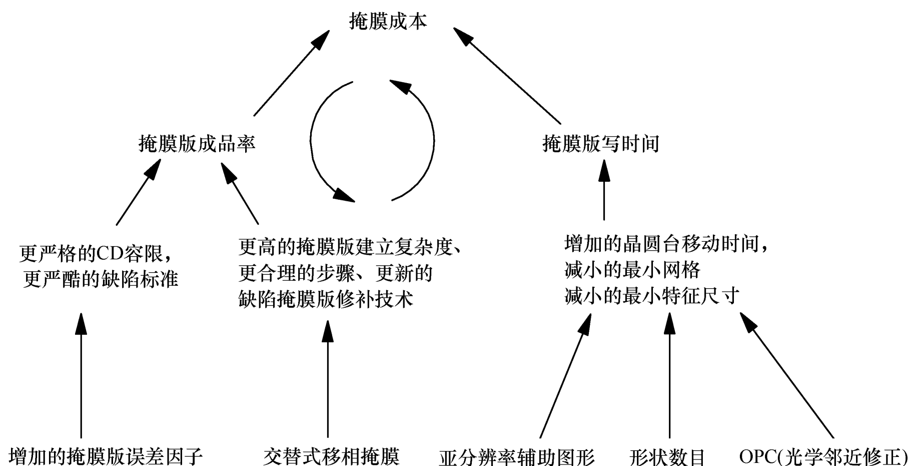


图 3-34　RET 与掩膜成本之间的关系

[各种 RET（交替式 PSM、辅助图形等）影响掩膜成品率和制备掩膜的写时间，从而影响掩膜版成本]

增强因素（*Mask Error Enhancement Factors*，*MEEF*）。因为 *CD* 容限随着工艺节点而缩放，掩膜对整体 CD 预算的贡献只有通过紧缩对掩膜 CD 的控制来维持。更严格的掩膜要求反过来又减小了掩膜成品率。

从缺陷的角度来看，掩膜制造者面临着两个挑战。在低 *k* 成像时代，对掩膜缺陷的敏感度增加了。另外，像交替式 PSM 这类掩膜的制备步骤有所增加，而增加的每个工艺步骤都会引入虽然不大但不可忽略的缺陷。此外，交替式 PSM 导致了较简单掩膜工艺中所没有的新类型缺陷（相位缺陷），并且通常很难修复。光学邻近修正以及亚分辨率辅助图形的使用均减小了最小特征尺寸，使得传递到掩膜写工具的数据量急剧增加，这样就急剧地增加了掩膜写时间。另外，较小的特征尺寸将使掩膜检查变得复杂。所有这些因素既增加了掩膜写时间，又降低了掩膜制备成品率，从而使掩膜成本上升。值得注意的是增加的掩膜写时间本身也将降低掩膜成品率，因为电子束曝光载物台的移动控制和稳定性要求将更加严格。

面对掩膜成本飞涨的困境，新设计的开发能力日益成为业界的关注焦点。业已提出了几个概念和方式来解决或至少缓解日益上涨的掩膜成本问题。降低有效成本的较明显方式之一是多项目晶圆的概念[38]。芯片制造商通过将多个设计组合在一组掩膜上的方法降低了每个设计的掩膜成本。实际上，使用同一组掩膜版曝光的晶圆总数得到增加，基本等于对每个设计单独曝光的晶圆的总和。每一种设计分别通过一些必要的数据准备和 OPC 操作进行处理，在生成最终数据时再合并在一起。这就缓解了接收和处理多个设计时的时间安排问题。虽然多项目晶圆似乎是降低掩膜成本的理想方式，但是还有几个问题值得注意。实际上只有当不同设计对应的晶圆工艺步骤，包括所有的金属层，都完全一致时才有可能有效使用这种技术。否则，相当部分的晶圆将会无用，这就冲淡了这一技术的优点，也与采用这一技术的初衷背道而驰。如果每种芯片设计符合标准尺寸，则划片过程中使得完好芯片受到损失的数目降至最小，更加有利于这一技术的有效使用。

减小成本上升影响的另一个有效方式是将定制芯片改变为现场可编程门阵列（FPGA）、可编程 ASIC、或在快速微处理器中通过软件实现一些需要采用互连实现的逻辑功能[39,40]。从芯片制造的角度来看，FPGA 具有只需要使用一组掩膜版的完全标准化了的晶圆加工的优势。最终用户要求的功能被编程进了单个芯片。遗憾的是，与采用 ASIC 实现的电路相比，它们的面积仍然大一到两个数量级，并且功耗明显增大。

综合 ASICs 和 FPGA 的特点，有一种结构化的 ASIC。对结构化 ASIC，在芯片制造商那里，使用一组掩膜版完成整个晶圆加工的前道工序，包括有源区、栅、一次接触，甚至实现早期金属层的工艺加工。然后将晶圆贮存在生产线上以待进一步的处理。按照功能要求通过顶层的两层或三层金属层将芯片上的单个逻辑单元连接起来，最终实现芯片功能。这样每个设计就只要承担最后三层金属的

掩膜费用，而其他层的费用则由不同芯片设计分摊。

3.5　发展前景：未来的光刻技术

回顾光刻的历史，可以明显地看到，要获得更小的尺寸，实际上是或多或少地沿着下述演化路径：

- 增加数值孔径
- 随着更多地使用 RET 而实现更小 k 因子成像，以及更严格的工艺波动控制
- 曝光波长的逐渐减小

尽管波长的变化可能干扰稳定发展的流程，但是，在过去这些发展还是相当成功，包括给技术带来极具生命力的光刻胶材料、透镜和掩膜版所用的光学材料的相应研究与革新。因为这些方式已经被证明非常成功，因此目前的趋势还是尽可能继续沿着这条道路前进，并且只要考虑到数十亿美元的需求也会为克服这些障碍提供足够的动力。试图采用不同的曝光技术的许多努力就像指引着一辆满载的火车沿着一条新的线路全速前进。采用不同的曝光技术如 x 射线和电子束光刻已经有很长的历史了。令人遗憾的是，迄今为止，光刻技术的发展还是跟不上微电路产品迅速发展的要求。

3.5.1　发展之路：157nm 光刻

波长小于 193nm 的激光光源，包括曝光波长为 157nm 的 F_2 激光与可能的 126nm 发射，以及主要曝光设备供应商正在开发的 157nm 曝光设备产品。除了光刻胶材料的实时可用性和激光源的可靠性等一般问题外，还有几个与 157nm 光刻问题相关的独特问题：与光学材料、污染以及用于保护掩膜版不受外部材料影响的聚合材料的可用性等相关的问题[41,42]。

对 157nm 波长，目前只有效地开发出两种吸收能力足够低的实用光学材料：氟化钙（CaF_2）和熔融石英（一种无定形的，高纯的石英结构）。现在的 248nm，甚至 193nm 系统中大部分的光学单元都采用熔融石英。熔融石英工艺技术已十分先进，材料的光学特性是完全各向同性的并其热膨胀系数极低。目前提得最多的熔融石英的替代物是氟化物，在现有的 193nm 曝光设备中就到了使用 CaF_2——特别是在那些要承受最高紫外线曝光密度的部件。与熔融石英相比，它优良的阻挡辐射损伤的能力使之成为这些情况下选用的材料。存在两种不同的光学材料供选用，也使得可以对与波长相关的焦点偏移进行修正。由于已在 193nm 系统中得到应用，较大 CaF_2 晶锭的晶体生长和抛光技术已经非常先进。令人遗憾的是，业已发现的材料光学特性（本征双折射）并不完全是

各向同性的，并且对 157nm 的影响远大于对 193nm[43]。这就对透镜设计提出了一个巨大的挑战，要求优化透镜材料方向，使得采用的材料将双折射效应减到最小[44]。可以采用激光，但是它的带宽太大。较窄的带宽非常关键，因为几乎没有哪种材料可以被用来修正像差。为了使较宽的激光线宽的影响最小，正在研究几种利用反射表面的其他透镜设计[45]，一个这样的设计是兼有反射光及折射光的系统。

对这一技术的进一步挑战与污染问题相关[46]。在 157nm 辐射下，用于清除透镜的碳氢化合物材料会残留在玻璃上，发生分解并在透镜表面形成沉淀物。这些沉淀物增加了透镜单元的吸收性。业已发现，氘灯发出的紫外光可以用来清除碳；碳被氧化成气体 CO_2 后挥发。然而，由硅分解形成的沉积物则很棘手并且难以清除。因此，选用那些不能被气化除去的材料时必须格外小心，并且精心设计气体的清除速率，使得残存的污染物保持在 ppt（parts per trillion，万亿分之一等级）水平。

采用 157nm 另外一个主要的技术挑战与掩膜相关：确切地说，是如何保护掩膜不受外界材料（Foreign Material，FM）影响。掩膜上的 FM 会在每次曝光中导致在每个晶圆上重复形成缺陷，在最坏情况下会导致芯片不能工作。为了保护掩膜不受污染物的影响，其上覆盖了一层与铬图形相距几毫米的有机薄膜。这个距离足以使缺陷处于焦点之外，从而小的颗粒不会引起印制缺陷；大部分较大的缺陷则可以通过机械方法除去。包括 193nm 的所有技术使用的是能够阻挡紫外光的损伤并且足够透明的聚合材料薄膜[47]。遗憾的是，虽然做了大量的研究，但是还没有发现一种适合于 157nm 工艺的材料。可以选用的替代物是熔融石英形成的保护性覆盖层。它们变得更厚，并在重力的作用下会产生弯曲，而这将影响图像的布局。

最可能使用 157nm 曝光设备的技术是 65nm 和 45nm 工艺。超高 NA 的 193nm 设备与功能强的 RET 的联合使用是这个节点基本光刻方法的强劲竞争者。当然，这些设备的机会窗口很窄，但是在 65nm 的试制生产中可能使用 193nm 设备。因此，几个主要的半导体制造商最近在他们的路线图中已不考虑 157nm 设备。

3.5.2 进一步演化：浸没式光刻

有时，仅通过重新考虑基本假设也可能实现重大突破。在寻求一项新光刻技术的过程中，在重新分析相关假设基础上提出的浸没式光刻技术是一个典型实例。重新分析描述一个系统分辨率能力的简单方程所包含的假设可知，该简单方程仅在空气情况下是精确的（即在折射率为 1 的材料中）。对于一般情况，应该把折射率 n 考虑进去，这时可以达到的分辨率 R 是[48]：

$$R = k\frac{\lambda}{\mathrm{NA} \cdot n}$$

式中，NA 是数值孔径，λ 是光源波长；k 是 k 比例因子。由于液体的折射率明显不等于 1，这个方法开启了具有巨大潜力提高分辨率的新途径。例如，对 193nm 波长，水的折射率是 1.44，这样在水中使用 193nm 光刻成像相当于在空气中使用 134nm 光刻，这对 193nm 的能力来说是个巨大的改善，而且还优于下一个可用的波长 157nm。由 B. Lin 提出的这种光刻工艺，称为浸没式光刻。这种类型光刻的示意图如图 3-35 所示。这种技术的主要特点是在晶圆台与透镜之间有液体，要考虑的主要问题是液体媒质不仅要具有高折射率，而且黏滞度也要较高，能保证晶圆台相对于透镜的正常移动[49,50]。按照晶圆台浸入液体的深度不同，现在已经提出了几种不同的方案[52]。从晶圆台完全浸入到液体中到仅在透镜下维持一个液体气泡。第一种方式被称作潜入式晶圆台，整个晶圆台在液体中移动。这一情况下可能需要添加一台泵，使液体流速度能适应要求的晶圆台移动速度。另一个代替方案是使晶圆台位于液体之外，而将晶圆放入位于晶圆台上一个装有液体的槽中。还有一种方案是建议仅在透镜下方存在液体槽，只有少量液

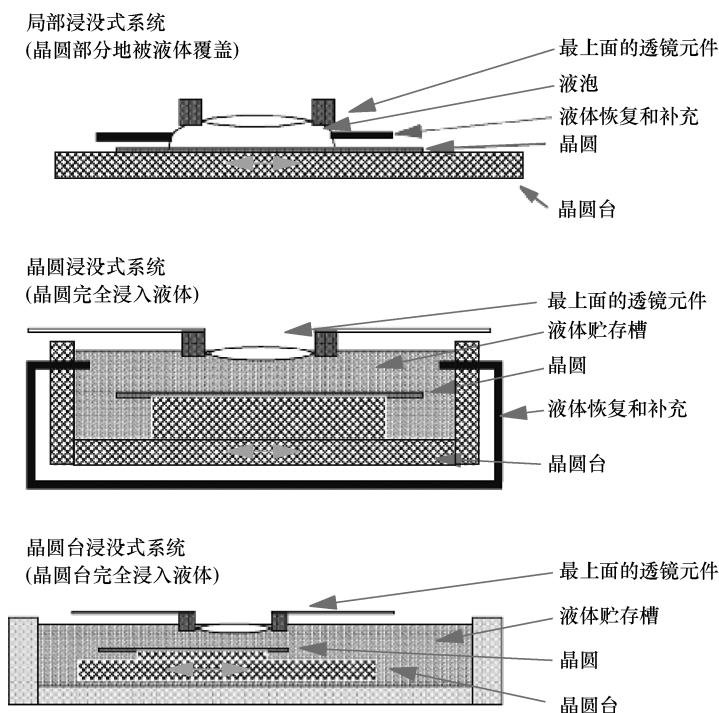


图 3-35　浸没式光刻示意图

（图中所示的三种不同方案，从上到下，浸没程度增加）

体随着透镜在晶圆上运动。与浸没式光刻相关的大部分技术问题是因为透镜与晶圆之间的高速移动，而位于他们之间的液体一定不能影响这种高速运动。特别是，在这种高速运动中，气泡的形成是主要问题。其他问题包括曝光时液体的加热问题和折射率的相关变化，能在水中成像的光刻胶的可用性是实施该技术的另一个潜在障碍。

这个方式只是最近才得到极大注意，曝光设备供应商对此展开了认真研究。这一技术的倡导者预计在 65nm 工艺中将会采用这一技术，但是这还是一个很大的挑战，因为目前这一曝光新技术还不成熟。

3.5.3 巨大突破：EUV 光刻

在曝光波长方面的进展是使用了超（极远）紫外线（extreme ultraviolet，EUV）成像，这不是很小的变化，而是成数量级的改进。为了评价完成这个任务所作的努力，我们应该注意到，从采用光刻工艺以来，曝光波长从 456nm 减少到 157nm，仅降低了 3 倍。与此相比，采用 EUV 应该是一项巨大的任务。采用这一技术所涉及的方方面面并不是简单的改进，而是巨大的突破，因为光源、曝光设备、掩膜、掩膜检查、透镜度量和光刻胶都与以前所使用的不同[53,54]。

EUV 成像采用的波长在 11 ~ 14nm 的范围内，之所以选择这一范围波长是因为目前已有的反射镜用于这一波长范围的光。这个波长范围位于紫外线与 x 射线之间的过渡区，因此有时也称为软 x 射线光刻。由于这个波长的光子具有巨大的能量，所有材料，包括气体，都对其起到吸收器的作用。因此不存在对其表现为透明的光学单元，光刻胶必须非常薄，并且光源不能使用固态罩。折射率相当小，这意味着单个界面的反射率较小。因此，采用透镜组也是有效的。在 11 ~ 14nm 的波长范围，最具潜力的光源是高温等离子材料，如氙。将高能量脉冲激光（Nd：YAG 激光）聚焦在液态或固态靶上就可以获得这些等离子[57]。在高强度激光照射下，材料汽化并在很短的时间内形成极高的温度（200000K）、相当高的密度（10^{17} ~ 10^{22}个电子/cm^2）、能发射具有期望波长的软 x 射线的少量等离子体。除了希望的辐射外，同时发射出的其他波长高强度光则必须滤去。因为采用的靶材料不断挥发，因此需要不断对靶材进行补充和替换。这些材料的最严重的缺点之一是固态靶材料形成的残存物。因为目前尚没有一种合适的透明材料用于保护反射镜，使得收集反射镜得到保护不受残存淀积物的影响就很困难。因此，冷冻的惰性气体如氙气就成为最具前途的靶材料[58]，即使从光发射的角度来看，固态材料如锡更加有效。另一种形成等离子体的方法是高度局部化的电脉冲放电[59]。然而，无论那种源，与达到要求的产量所需的目标值相比，源效率还差约一个数量级。

所有的成像必须使用反射镜完成。由于较低的折射率，单层反射率较小，因

此反射镜将采用多层系统。多层系统中大量的界面使总的反射率得到提高。这种结构中，尽管单层的反射光很少，许多反射层的叠加作用将产生相当大的反射率。单层的厚度必须控制在一定范围，使得每个界面反射的光束之间发生明显干涉。每层厚度约为几个纳米的级别，并且厚度波动范围需要控制在零点几个纳米之内。因为这些反射镜通过相长干涉实现需要的反射率，反射率与入射角关系非常密切，厚度必须修正。多层反射镜由 Si 和 Mo 层交替重叠组成，由于它们的原子质量不同，折射率也不同[60]。

与该技术相关的其他光学单元类似，需要采用反射掩膜而不是透明掩膜[61]。EUV 掩膜是由位于多层反射镜顶部的印有光刻图形的 EUV 吸收材料构成。这种掩膜存在的最严重的问题与多层系统中的缺陷有关，而目前对于这类缺陷还没有修补技术。

目前已经可以采用折射率相当低的反射镜，由 EUV 工具形成光学版图的主要目标是尽量减少光学单元的数目。图 3-36 是一个使用六个界面的系统[53]示意图。能对相当大面积成像的反射系统所固有的版图复杂性使得数值孔径被限制到 0.3[62]，由于与现有系统的波长有极大差异，这并不是一个重要的问题。

目前已经开发出了一个 EUV 工程测试台，可以在数值孔径为 0.1 情况下实现大范围（26mm×32mm）曝光[63]。目前 ASML 和一个包括 Nikon 与 Canon 的日本团队正在开发这类系统，估计价格约为五千万美元。

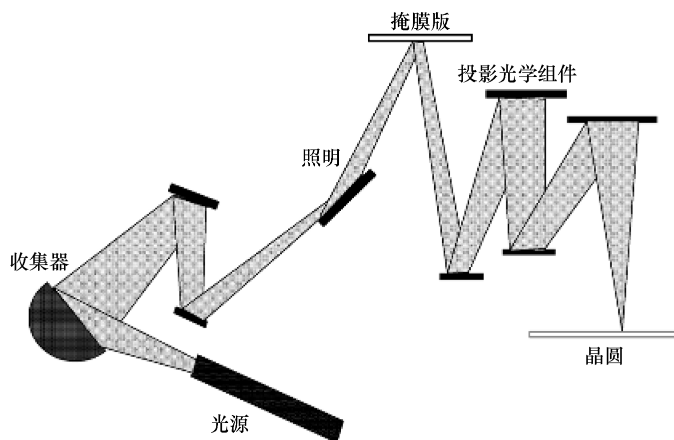


图 3-36　EUV 光学系统示意图

3.5.4　粒子束光刻

使用带电粒子（电子或离子）光刻是可供下一代光刻技术选用的另一种方法。一般地说，这些系统可以按照采用的写策略（直接写或者投影）和带电粒子的类型（离子或者电子）加以分类。由于运动粒子的波长与其质量和速度有

关，因此粒子束光刻设备的分辨率也会受到其制约。幸运的是，对应的波长非常小，使用电子束可以实现 10nm 级别的分辨率。

3.5.5 直写式电子束设备

直写式电子束设备在掩膜制造中已经使用了很长时间[65,66]。图 3-37 是这种系统的示意图。从源发射的一束电子通过一组磁透镜聚焦在晶圆上。由于电子束源的尺寸有限，必须在最后可获得的光点直径与电子束强度之间进行折衷。虽然最后的光点直径很大程度上是由电子源的尺寸决定的，但是采用高束流可以提高产量，为此要求大面积的电子源。最初采用的方法是加热硼化镧顶端的电子束源。经由顶端的电子发射是通过加热以及在顶端覆盖一种低功函数材料（即对电子发射的能量势垒较小）实现的。然而，对较高的分辨率系统来说，这些尖端的亮度还不够，因此已经被场致发射源所取代[67]。在这些源中，金属顶部尖端处集中的电场将电子从金属顶部牵引出。这种源对尖端只要稍许或者不需要加热就能发射电子，这样就减少了发射电子的动能分散。对于存在高阶色差的电子光学，这就有利于提高成像能力。电子光学的基础是磁场以及磁场

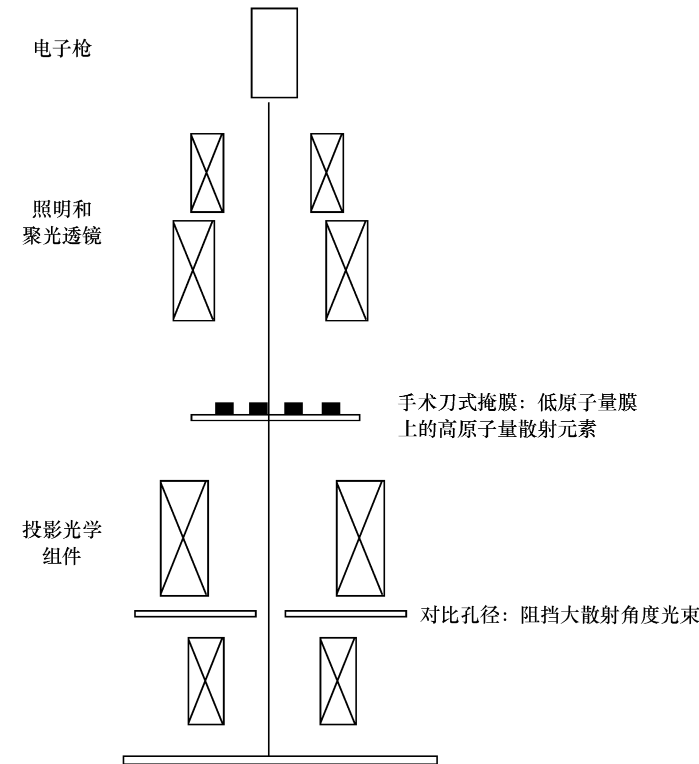


图 3-37 电子束投影设备示意图

对带电粒子的作用力。与光学透镜相似，具有适当形状磁极的电磁铁可以对分散的电子束起聚焦作用[68]。铁磁体材料能使磁场集中并有助于得到需要的磁场分布。在用于电子束写入设备的情况，它们的磁性需要被软化，这样在需要修正磁场强度时，它们不会维持在高磁场状态。遗憾的是，从像差的角度来看，即使是靠近系统轴线移动的电子，电子光学的质量也相当差，而对于与轴成较大的角度移动的电子束，光学质量将急剧退化。这就限制了这些透镜的最大数值孔径，并限制了可实现的最大图像尺寸。电子束还存在一种光学成像情况并不存在的图像模糊机制[69]。由于电子的相互排斥，聚焦在较小区域的电子束呈现出的随着电流增加而更加严重的图像模糊。这一机制限制了高束流时的成像能力，而从产量的角度来说高束流是必须的。

采用静电偏转器或磁偏转器可以实现束流偏转[70]，与普通的电视显像管中使用的相似。与不需要纳米级光斑的电视显象管不同，电子束曝光设备中电子束的偏转距离相对较小（约几十个微米），这是因为电子色差随着离开轴线而快速增加。因此，如果晶圆台不移动的话，电子束能能够写的范围较小。图形以光栅的方式写入，与电视屏幕上的图像相似。电子束通过对一条互连的空间信息控制电子束的开-关信息编码，完成对线条的扫描。扫描完一条线条后，继续写下一条线条。在完成约 50μm×50μm 的面积扫描后，晶圆台移动到下一个方形区域的中心，继续扫描。

在光刻中采用这一技术的主要困难在于数据是串行传送而不是如成像系统中那样采用并行方式。即使通过努力，采用并行电子束方法增加数据传送速率，也改变不了生产效率低下这一恼火的问题，例如，关键掩膜的曝光时间是以小时计。电子束曝光的其他问题与下述事实相关：电子束会导致衬底带电。如果所带的电不能有效的去除，会导致不希望的束偏转。在先进工艺中，栅氧厚度越来越薄，而高能电子束会导致栅氧损伤。此外，与 EUV 系统类似，电子束曝光系统是在真空环境下工作，很难实现晶圆的快速装载，这也增加了系统的复杂程度。

电子束直写系统的最大优势之一是不需要采用昂贵的掩膜版，因此这个技术对生产量很低而掩膜成本非常昂贵的 ASIC 设计者来说极具吸引力。生产效率低的缺点对图形密度较低的层次（如接触层）来说不成为问题。当前市场上已实现商品化的可用系统并不多。目前这一领域最令人兴奋的发展[71]是利用在微机械系统（MEMS）领域发展的制造能力，尝试使用电子束源阵列。现在正研究的方案是采用 64×64 的源阵列并对每个源都有独立偏转控制，可以实现 250μm×250μm 的场覆盖。有待进一步解决的问题是能否基于这一技术构建均匀性足够好并且寿命足够长的电子源，进而组建可靠的电子束曝光工具。

生产效率不高的问题可以结合采用光刻中的投影技术来解决。该方案的要点

是，不再逐个地写单个像素，而是照射较小的区域（亚场）并将其投影到晶圆上[72]。单个亚场整合成完整的曝光场。即使亚场之间用支撑件机械隔开，电子束系统的精确偏转控制也可以实现亚场投影[73]，通过整合工艺可以完全消除亚场之间未被曝光的区域。这一能力非常重要，使得保证掩膜机械稳定性成为可能。因为一方面为了使电子束能够通过，掩膜必须足够薄。为了使掩膜具有足够的机械稳定性，则可以允许在单个亚场之间放置支撑件支柱。

为了提高曝光设备的生产效率，需要采用高束流电子束，由此引起的掩膜加热是制约这一技术的效应之一。由 Bell 实验室开发的 SCALPEL 系统是解决这一问题的重大突破[73]。其关键是改变了原先在掩膜上采用吸收元素的方式，使电子束被衍射并且衍射的电子束由位于设备内部的光栅吸收，这样把所产生的热的吸收转给了对成像过程不重要的设备单元。衍射是采用高原子质量的材料完成的。

虽然投影概念极大地提高了系统的生产效率，但是仍然存在的一个重要制约是电子束的偏转能力有限，这就要求晶圆台的大范围移动。这个问题已在 IBM 开发的采用可变轴浸入式透镜减小曝光的 PREVAIL 系统中得到解决。在这种特殊的设计中，能形成高分辨率成像所必须的最小色差透镜的光学中心可以被横向移动，为此只要在聚焦透镜场上叠加附加的偏转场。采用这一技术可以极大改善电子束的偏转，在掩膜版上成像范围高达 10mm，在晶圆平面上达 ±2. 5mm。

当初用于 SCALPEL 系统的技术被转移到了 ASML 和应用材料联合开发的商业系统中。遗憾的是，这个叫 eLith 的联合企业，在 2001 年初解散了[75]。然而，仍然有一个主要的曝光设备供应商准备提供商用电子束投影设备[76]。

离子束投影光刻[77]是被讨论的最后一项技术。与电子束直写设备相似，离子束已经在掩膜修补和 300mm 晶圆的非损坏性微表征中找到了用武之地。在这些领域，利用了离子束的淀积和去除材料的能力。由奥地利团队开发的系统是投影光刻设备。直写方式最近已经面世，并且在磁性存储碟片中得到应用。

离子束投影光刻系统由三个主要部分组成：源、柱和放置晶圆的腔体。用于掩膜修补的聚焦离子束系统中常用的是镓离子，而在光刻设备中使用带电的氢或氦原子。在这些源中，采用的材料以气态形式引入并经电子轰击发生电离。设置合适的电场，从源中抽取出离子并将其注入到电子光学柱中。这个系统使用的掩膜与电子束投影系统中相似，是镂空掩膜。不像大多数电子束系统那样使用磁性光学透镜，离子束穿过掩膜版后，进入一个多电极静电透镜系统，聚焦在晶圆台上。

参 考 文 献

[1] S. Okazaki, Lithography for ULSI, *Proc. SPIE*, Vol. 2440, pp. 18–32, 1995.

[2] H. Levenson, *Principles of Lithography*, SPIE Press, Bellingham, WA, 2001.

[3] R. Dejule, Resist enhancement with antireflective coating, *Semicond. Int.*, July 1996, p. 169.

[4] Y. Borodovsky, Impact of local partial coherence variation on exposure tool performance, *Proc. SPIE*, Vol. 2440, pp. 750–770, 1995.

[5] J. Kirk, Review of photo resist based lens evaluation methods, *Proc. SPIE*, Vol. 4000, pp. 2–8, 2000.

[6] P. Dirksen, C. Juffermans, A. Engelen, P. deBisschops, and H. Muellerke, Impact of high order aberrations on the performance of the aberration monitor, *Proc. SPIE*, Vol. 4000, pp. 9–17, 2000.

[7] N. Farrar, A. Smith, D. Busath, and D. Taitano, Measurement of lens aberration by in-situ interferometer and classification for correct application, *Proc. SPIE*, Vol. 4000, pp. 18–29, 2000.

[8] C. Progler and A. Wong, Zernike coefficients, are they really enough, *Proc. SPIE*, Vol. 4000, pp. 40–52, 2000.

[9] D. Flagello, J. Mulkens, and C. Wagner, Optical lithography into the millennium sensitivity to aberrations, vibrations and polarization, *Proc. SPIE*, Vol. 4000, pp. 172–183, 2000.

[10] M. Maenhoudt, S. Verhaegen, K. Ronse, P. Zandbergen, and E. Murzio, Limits of optical lithography, *Proc. SPIE*, Vol. 4000, pp. 373–387, 2000.

[11] A.K. K. Wit, *Resolution Enhancement Techniques*, SPIE Press, Bellingham, WA, 2001.

[12] W. N. Partlo, P. J. Thompkins, P. G. Dewa, and P. F. Michaloski, Depth of focus and resolution enhancement for i-line and DUV lithography using annular illumination, *Proc. SPIE*, Vol. 1972, pp. 753–764, 1993.

[13] K. Tounai, H. Tanabe, H. Nozue, and K. Kasama, Resolution enhancement with annular illumination, *Proc. SPIE*, Vol. 1674, pp. 753–764, 1992.

[14] Y. Granik, N. Cobb, and T. Do, Universal process modeling with VTRE for OPC, *Proc. SPIE*, Vol. 4691, pp. 377–394, 2002.

[15] K. D. Lucas, J. C. Word, G. N. Vandenberghe, S. Verhaegen, and R. M. Jonckheere, Model-based OPC for first generation 193 nm lithography, *Proc. SPIE*, Vol. 4346, pp. 119–130, 2002.

[16] O. Otto and R. Henderson, Advances in process matching for rules-based optical proximity correction, *Proc. SPIE*, Vol. 2884, pp. 323–332, 1996.

[17] F. X. Zach, D. J. Samuels, A. C. Thomas, and S. A. Butt, Process dependencies of optical proximity corrections, *Proc. SPIE*, Vol. 4346, pp. 113–118, 2001.

[18] J. P. Stirniman and M. L. Rieger, Spatial filter models to describe IC lithographic performance, *Proc. SPIE*, Vol. 3051, pp. 469–478, 1997.

[19] Y. Granik, Correction for etch proximity: new models and applications, *Proc. SPIE*, Vol. 4346, pp. 98–112, 2001.

[20] J. Garofalo, O. Otto, R. Cirelli, R. Kostelak, and S. Vaidya, Mask assisted off axis illumination technique for random logic, *J. Vac. Sci. Technol. B*, Vol. 11, pp. 2651–2658, 1993.

[21] J. Garofalo, O. Otto, R. Cirelli, R. Kostelak, and S. Vaidya, Automated layout of a mask-assisted for realizing 0.5 k1 ASIC lithography, *Proc. SPIE*, Vol. 2440, pp. 302–312, 1995.

[22] A. H. Gabor et al., Subresolution assist feature implementation for high performance logic gate level lithography, *Proc. SPIE*, Vol. 4346, pp. 418–425, 2002.

[23] M. Levenson, N. Viswanathan, and R. Simpson, Improving resolution in photolithography with a phase shift mask, *IEEE Trans. Electron Devices*, Vol. 29, pp. 1812–1846, 1982.

[24] G. Mack, Fundamental issues in phase shifting mask technology, *OCG Microelectronics Conference*, pp. 23–25, 1993.

[25] K. Adam and A. R. Neureuther, Algorithmic implementation of domain decomposition methods for the diffraction simulation of advanced photomasks, *Proc. SPIE*, Vol. 4691, pp. 107–124, 2002.

[26] L. Liebmann, N. A. Northrup, J. Culp, L. Segal, A. Barish, and C. Fonseca, Layout optimization at the pinnacle of optical lithography, *Proc. SPIE*, Vol. 5042, pp. 1–14, 2003.

[27] J. A. Torres, F. Schellenberg, and O. Toublan, Model assisted double dipole decomposition, Vol. 4691, p. 407–417, *Proc. SPIE*, 2002.

[28] S. M. Mansfield, L. Liebmann, A. Molless, and A. Wong, Lithographic comparison of assist feature design strategies, *Proc. SPIE*, Vol. 4000, pp. 63–76, 2000.

[29] R. Socha, M. Dusa, L. Capodieci, J. Finders, F. Chen, D. Flagello, and K. Cummings, Forbidden pitches for 130 nm lithography and below, *Proc. SPIE*, Vol. 4000, p. 1140, 2000.

[30] L. W. Liebmann et al., Optimizing style options for subresolution assist features, *Proc. SPIE*, Vol. 4346, pp. 141–145, 2001.

[31] L. Liebmann, J. Lund, F. L. Hueng, and I. Graur, Enabling alternating phase shift mask designs for a full logic gate level: design rules and design rule checking, *Proc. DAC*, pp. 79–84, 2001.

[32] M. L. Rieger, J. P. Mayhew, and S. Panchapakesan, Layout design methodologies for sub-wavelength manufacturing, *Proc. DAC*, pp. 85–88, 2001.

[33] M. Sanie, M. Cote, P. Hurat, and V. Malhotra, Practical application of full feature alternating phase shift technology for a phase aware standard-cell design flow, *Proc. DAC*, pp. 93–96, 2001.

[34] U. Behringer, Foreword, EMC2003 (European Mask Conference), p. 1, Jan. 13–15, 2003.

[35] A. Balasinski, Mask cost for sub-100 nm technologies: stopping a runaway? *Proc. SPIE*, Vol. 5043, pp. 82–92, 2003.

[36] Lithography CoO analysis, *http://www.sematech.org*.

[37] D. Pramanik, H. Kamberian, C. Progler, M. Sanie, and D. Pinto, Cost effective strategies for ASIC's masks, *Proc. SPIE*, Vol. 5043, pp. 142–148, 2003.

[38] R. D. Morse, *Proc. SPIE*, Vol. 5043, pp. 100–112, 2003.

[39] G. Prophet, Structured ASICs: more gain, less pain? *EDN Europe*, Aug. 7, 2003; *http://www.reed.electronics.com/electronicnews/community/22113/Semiconductors?starting=651*.

[40] B. Dipert, Silicon segmentation, *EDN*, Sept. 18, 2003; *http://www.reed-electronics.com/electronicnews/article/CA321801?stt=000&industryid=22113&industry=Semiconductors*.

[41] S. Dana, Pushing the limits, *OE Mag.*, pp. 20–22, Mar. 2002; *http://www.oemagazine.com/fromTheMagazine/mar02/pushingthelimits.html*.

[42] S. Dana, Progress report: 157 nm lithography prepares to graduate, *OE Mag.*, Feb. 2003; *http://www.oemagazine.com/fromTheMagazine/feb03/157.html*.

[43] J. Burnett et al., *Phys. Rev. B*, Vol. 64, No. 241102, 2001.

[44] W. Ulrich, S. Beiersdorfer, and H. J. Mann, Trends in optical design of projection lenses for UV and EUV lithography, *Proc. SPIE*, Vol. 4146, pp. 13–24, 2000.

[45] J. Burnett, Intrinsic birefringence in calcium fluoride forces optical engineers to use sophisticated design techniques for 157 nm lithography systems, *OE Mag.*, Mar. 2002; *http://www.oemagazine.com/fromTheMagazine/mar02/bire.html*.

[46] R. R. Kunz, V. Liberman, and D. K. Downs, Experimentation and modeling of organic photocontamination on lithographic lenses, *Proc. SPIE*, Vol. 4000, p. 474, 2000.

[47] R. H. French et al., 157-nm pellicles: polymer design for transparency and lifetime, *Proc. SPIE*, Vol. 4691, pp. 576–583, 2002.

[48] B. J. Lin, New λ/NA scaling equations for resolution and depth of focus, *Proc. SPIE*, Vol. 4000, pp. 759–764, 2000.

[49] J. Hoffnagle, W. D. Hinsberg, M. Sanchez, and F. A. Houle, *J. Vac. Sci. Technol. B*, Vol. 17, p. 3306, 1999.

[50] B. Smith, H. Kang, A. Burov, F. Cropranese, and Y. Fen, Water immersion lithography for the 45 nm node, *Proc. SPIE*, Vol. 5040, pp. 679–699, 2003.

[51] Immersion lithography: successor to super high NA 193 nm technology? *http://eedesign.com/pressreleases/bizwire/111879*.

[52] S. Owa and H. Nagasaka, Immersion lithography: its potential performance and issues, *Proc. SPIE*, Vol. 5040, pp. 724–733, 2003.

[53] C. Gwyn and P. Silverman, EUV transition from research to commercialization, *Proc. SPIE*, Vol. 5130, pp. 990–1004, 2003.

[54] N. Harned and S. Roux, Progress report: engineers take the EUV lithography challenge, *OE Mag.*, Feb. 2003; *http://www.oemagazine.com/fromTheMagazine/feb03/euv.html*.

[55] B. Lai and F. Cerrina, Image formation in multilayer optics: the Schwartzschild objective, *Proc. SPIE*, Vol. 563, pp. 174–179, 1985.

[56] T. E. Jewell, J. M. Rodgers, and K. P. Thompson, Reflective systems design for soft x-ray projection lithography, *J. Vac. Sci. Technol. B*, Vol. 8, pp. 1509–1513, 1990.

[57] U. Stamm et al., *Proc. SPIE*, Vol. 4688, pp. 122–133, 2002.

[58] B. A. Hansson, L. Rymell, M. Berglund, O. E. Hemberg, E. Janin, J. Thoresen, S. Mosesson, J. Wallin, and H. M. Hertz, Status of the liquid-xenon-jet laser-plasma source for EUV lithography, *Proc. SPIE*, Vol. 4688, pp. 102–109, 2002.

[59] N. Fornaciari et al., Power scale-up of the extreme-ultraviolet electric capillary discharge source, *Proc. SPIE*, Vol. 4688, pp. 110–121, 2002.

[60] E. Luis et al., Progress in Mo/Si multilayer coating technology for EUV mirrors, *Proc. SPIE*, Vol. 3997, pp. 406–411, 2000.

[61] P. J. Mangat, S. D. Hector, S. Rose, G. F. Cardinale, E. Tejnil, and A. R. Stivers, EUV mask fabrication with Cr absorber, *Proc. SPIE*, Vol. 3997, pp. 76–82, 2000.

[62] T. Oshino et al., Development of illumination optics and projection optics for high NA EUV exposure tool, *Proc. SPIE*, Vol. 5037, pp. 75–82, 2003.

[63] D. J. O'Connel et al., Lithographic characterization of improved projection optics in the EUVL engineering test stand, *Proc. SPIE*, Vol. 5037, pp. 83–94, 2003.

[64] H. Meiling et al., The EUV program at ASML: an update, *Proc. SPIE*, Vol. 5037, pp. 83–94, 2003.

[65] D. R. Herriot, R. J. Collier, D. S. Alles, and J. W. Stafford, EBES: a practical electron lithography system, *IEEE Trans. Electron Devices*, Vol. 22, pp 385–392, 1972.

[66] F. Abboud, J. Poreda, and R. L. Smith, MEBES IV: a new generation raster scan electron beam lithography system, *Proc. SPIE*, Vol. 1672, pp. 111–125, 1992.

[67] M. Gesley, MEBES IV thermal field emission tandem optics for electron beam lithography, *J. Vac. Sci. Technol. B*, Vol. 9, pp. 2949–2951, 1991.

[68] P. W. Hawkes, *Selected Papers on Electron Optics*, SPIE Press, Bellingham, WA, 1994.

[69] L. R. Harriot et al., Space charge effects in projection charged particle lithography systems, *J. Vac. Sci. Technol. B*, Vol. 13, pp. 2402–2408, 1995.

[70] L. H. Lin and H. L. Beauchamp, High speed beam deflection and blanking for electron lithography, *J. Vac. Sci. Technol.*, Vol. 10, pp. 987–990, 1973.

[71] P. Ware, Removing the mask, *OE Mag.*, pp. 26–27, Mar. 2002.

[72] A. E. Novembre et al., Fabrication and commercialization of scalpel masks, *Proc. SPIE*, Vol. 3412, pp. 350–357, 1998.

[73] J. A. Liddle, L. R. Harriot, A. E. Novembre, and W. K. Waskiewicz, SCALPEL: a projection electron beam approach to sub-optical lithography, *http://www1.bell-labs.com/project/SCALPEL/*.

[74] R. S. Dhaliwal et al., PREVAIL: electron projection technology approach for next-generation lithography, *IBM J. Res. Dev.*, Vol. 45, pp. 615–638, 2001.

[75] D. Lammers, EUV gains as venture ends e-beam litho work, *EETimes*, Jan. 2001; *http://www.eetimes.com/story/OEG20010105S0023*.

[76] S. Fukui, H. Shimizu, W. Ren, S. Suzuki, and K. Okamoto, Nikon EB stepper: its system design and preliminary performance, *Proc. SPIE*, Vol. 5037, pp. 504–511, 2003.

[77] H. Loescher, Masked ion beam lithography and direct structuring of curved surfaces, *Proc. SPIE*, Vol. 5037, pp. 156–161, 2003.

第 4 章　混合信号电路设计

4.1　引言

基于更先进工艺（0.13μm 和小于 0.13μm）的混合信号设计提出了一个在初始开始体系架构的设计阶段以及在具体设计阶段都必须考虑的问题。目前许多新产品都把模拟电路和混合信号电路与较大的数字电路结合在一起。微处理器是这种趋势的典型实例，因为它们一般都包括锁相环（PLL）、频率合成器、延迟锁相环和数据转换器。由于这类芯片中高速数字逻辑占用了绝大部分芯片面积，因此在开发相应工艺时主要考虑如何解决与数字电路相关的问题，这种选定工艺的思路就迫使模拟和混合信号设计者需要采用特殊的设计方法来确保电路能满足严格的要求，而不是依靠专门的工艺。本章将概述在深亚微米工艺中成功设计模拟和混合信号电路必须考虑的主要问题，同时提出避免各种隐患的建议。

4.2　设计考虑

对于持续不断向更小几何尺寸发展的技术，在工艺方面已经取得了一些重要进展。这些进展包括极薄的栅氧（对 130nm 工艺，物理厚度或者等效氧化层厚度（EOT）范围为 1.5 ~ 2.2nm，而对 90nm 工艺，则为 1.1 ~ 1.7nm）、更浅的源和漏扩散、更浅的槽隔离、多种氧化层厚度、给定氧化层厚度时的多阈值电压、以及多种金属厚度。所有这些工艺进展已经促使了半导体市场的爆炸性增长，而代价是成功开发新产品所需的设计方法学和工具的日益复杂。由此，混合信号设计者将面临如下问题：

- 模型的精度
- 泄漏电流（包括源—漏极间泄漏与栅泄漏）
- OD 应力效应
- 工艺起伏变化的影响（包括器件与器件之间，芯片与芯片之间，晶圆与晶圆之间，批次与批次之间，以及厂商与厂商之间的起伏变化）
- 电源动态裕度效应
- 噪声隔离
- 功耗

本章将讨论上述问题，并给出针对性设计建议。

4.3 器件建模

针对下一代工艺的挑战性的设计目标，能够精确反映电路物理特性的器件模型就变得更加复杂。对 90nm 及尺寸更小的工艺，成功设计混合信号集成电路要求采用精度更高的器件模型。例如相对于 0.13μm，除了增加模型的电流参数外，还必须考虑几个在以前工艺中可以忽略的新因素，包括：

- 浅槽隔离（STI）应力感应效应
- 栅泄漏电流
- 栅感应漏电流（GIDL）
- 应力感应二极管漏电流
- 多晶硅栅耗尽效应
- 浅漏源的串联电阻
- 负偏温度不稳定性（NBTI）

BSIM3 模型已经考虑了上述很多问题，但是这些效应的影响随着尺寸的缩小变得更加显著。图 4-1 所示为一个包括了几个上述寄生元素的 MOS 器件截面简图。在先前的工艺中，其中许多元素因为与其他寄生元素（如扩散电容）相比

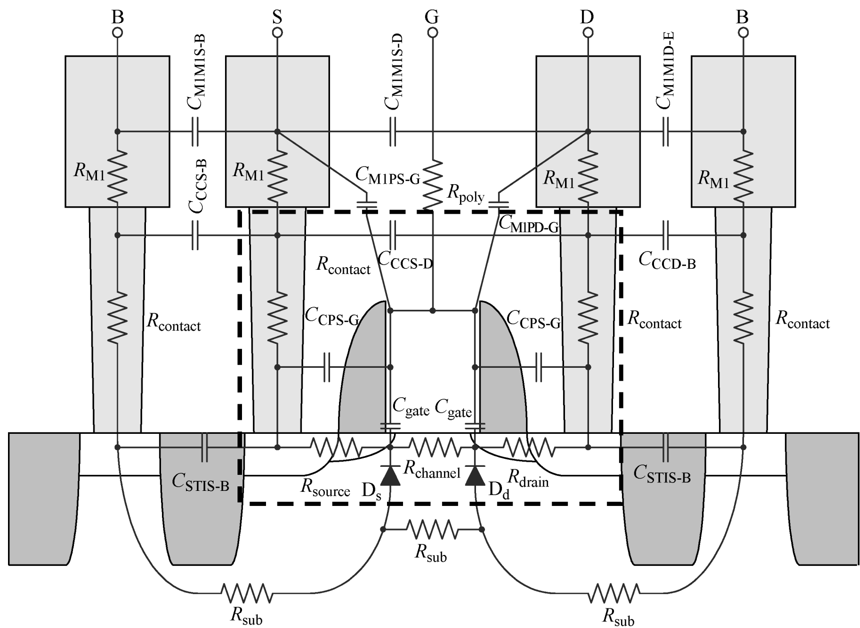


图 4-1 晶体管模型

较小而被忽略，但是对更先进的工艺，如 0.13μm 和小于 0.13μm 的工艺，它们将不能再被忽略了。建模过程的关键是要明确基本模型中要包括哪些元素，布局前仿真采用的子电路必须包括哪些元素，而在寄生参数提取部分又必须包括哪些元素。图 4-1 所示器件中，虚线区域代表基本器件模型包括的部分。为了在能够产生合理结果的同时，又能使模型的复杂度最小，关键问题是需要确定哪些参数对器件的性能影响最大。调整关键的设计规则能使一些寄生元素的影响最小，例如增加栅与接触孔的间距。这等效于对同一设计执行了两次设计规则检查，一次是针对设计的数字部分，而另一次则是针对设计的模拟部分。

表 4-1 总结了模型中通常考虑的各种寄生元件以及基于典型的 90nm 工艺的数值范围。最右边的一栏列出了相应数值的可能偏差，作为混合信号电路分析中是否需要考虑该寄生参数时参考。该列中的空白栏意味着很难给出其偏差范围，因为他们与特殊的应用有很大关系。

表 4-1　器件模型寄生参数和 90nm 典型工艺的取值范围

参数	范　围	引起波动的原因	可 否 忽 略	可能的误差
R_{M1}	0.05 ~ 0.2Ω/sq	腐蚀或凹陷与金属 CD 控制引起的金属厚度变化	局部布线（＜20μm）：典型情况可以忽略；长步线：不能忽略	局部布线 0 ~ 2%，长点的互连 2% ~40%
$R_{contact}$	10 ~ 25Ω/cnt	接触孔套刻误差接触孔刻蚀后的填充百分比	通常为基本模型的一部分	
R_{poly}	5 ~ 12Ω/sq	多晶厚度，金属硅化物波动，多晶 CD 控制	小宽度器件：可以忽略；大宽度器件：不能忽略；RF 应用：不能忽略	
$C_{M1M1S\text{-}B}$，$C_{M1M1D\text{-}B}$	0.1 ~ 0.3fF/μm	腐蚀、凹陷（影响程度较轻）以及金属 CD 控制引起的金属层厚度波动	源 - 衬连接时：可以忽略；漏与衬底连接时：不能忽略，但与版图关系非常密切	1% - 2%
$C_{M1M1S\text{-}D}$	0.05 ~ 0.1 fF/μm	金属厚度，金属 CD 控制	模型中必须包括	1% ~ 3%
$C_{M1PS\text{-}G}$，$C_{M1PD\text{-}G}$	0.02 ~ 0.06fF/μm	金属与多晶厚度，金属与多晶 CD 控制，介质（TEOS）厚度波动	模型中必须包括	1% ~ 3%
$C_{CCS\text{-}B}$，$C_{CCD\text{-}B}$，$C_{CCS\text{-}D}$	0.02 ~ 0.05fF/cnt	接触孔刻蚀	典型情况可以忽略	<2%

（续）

参数	范　　围	引起波动的原因	可 否 忽 略	可能的误差
$C_{CPS\text{-}G}$，$C_{CPD\text{-}G}$	0.02 ~ 0.07fF/cnt	接触孔刻蚀，多晶厚度，多晶 CD 控制	通常为基本模型的一部分	<1%
$C_{STIS\text{-}B}$，$C_{STID\text{-}B}$	0.03 ~ 0.07fF/μm	STI 宽度	可能被包括进模型中，与版图关系密切	<1%，除非存在邻近器件的耦合噪声
R_{sub}	500 ~ 1000Ω/sq	衬底掺杂	可能被包括进模型中，特别是在高速设计和进行噪声仿真时	
C_{gate}	约 12 fF/μm^2	氧化层厚度，工作电压	必须包括进模型（栅电容是集总值，但是实际上它由几部分组成）	

图 4-2 所示为版图设计后期布线寄生电阻设为 0 的一个模块中提取的结果与实际测量值之间的误差。从这个模块中测量了多达 72 个延迟时间，其中 2/3 的路径出现的误差小于 5%。这看起来相当小，但是如果考虑一个工作频率为 10GHz 的高速时钟数据恢复芯片，其时钟周期仅 100ps，因此 5% 的误差将是 5ps，这足以产生时序错误，特别是与其他定时误差源叠加时问题将更加严重。

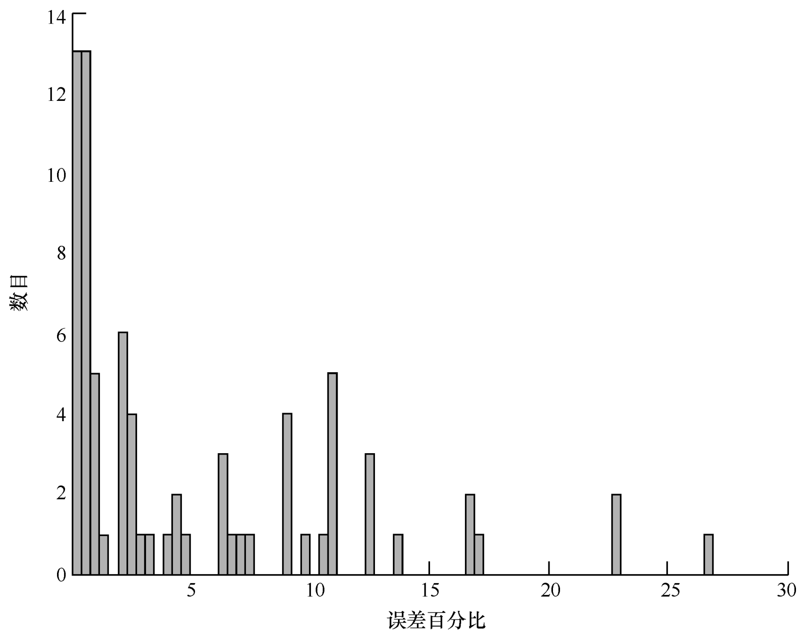


图 4-2　90nm 工艺情况下忽略金属和多晶电阻导致的误差

图 4-3a 所示为 90nm 工艺锁相环电路中高速数字部分的节点电容。大部分节点电容小于 5fF，但是有几个电容由于增加了互连寄生电容而变成了重负载。图 4-3b 中显示的是压控振荡器（VCO）单元节点电容的常用对数。x 轴的零点代表 1fF；由图可见大部分电容值小于 1fF，但是节点电容导致 VCO 单元振荡频率降低。附加效应包括不希望的与邻近节点的耦合，这将引起很大的相位噪声从而使性能退化。

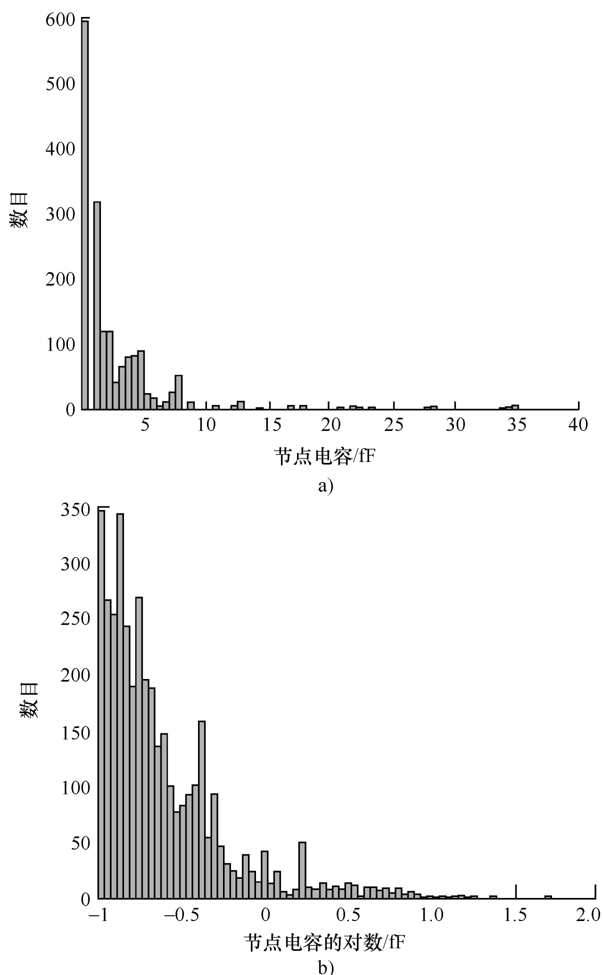


图 4-3　采用 90nm 工艺的锁相环电路中的节点电容

a）高速数字模块的节点电容　b）压控振荡器的节点电容

更新的 BSIM4 模型考虑了 STI 应力感应效应和栅泄漏效应。STI 应力效应的引入并不影响仿真时间，但是栅泄漏效应将使仿真时间急剧增加。栅泄漏效应与

偏置有关，因此在仿真运行期间每个工作点计算均需重新进行估算，而不是只要在仿真开始时计算一次即可。STI 应力效应与版图密切相关，这需要了解如何对一个电路进行版图设计，才能在版图设计前的仿真中正确考虑 STI 应力效应。该应力效应的发生是因为硅和氧化硅的热膨胀系数不同，将导致硅中产生压缩应力[1,37]。压缩应力会引起沟道中载流子迁移率的变化，如图 4-4 所示。对较小几何尺寸的器件，两个 STI 区域之间的间距更小，这个效应将变得更加显著。迁移率可以采用下式表征：

$$\mu_{\text{eff}} = \mu_{a\ \text{min}}\left(1 + \Delta\mu_{\text{max}}\frac{a - a_{\text{min}}}{a}\right)$$

式中，$1/a = (1/SA + 1/SB)$；a_{min}是工艺设计规则规定的扩散区之间最小间距；$\mu_{a\ \text{min}}$是在栅到扩散之间的间距为最小值时的迁移率；而$\Delta\mu_{\text{max}}$是相对于$\mu_{a\ \text{min}}$的迁移率最大变化值。对 PMOS 器件，$\Delta\mu_{\text{max}}$是负值，而对 NMOS 器件，$\Delta\mu_{\text{max}}$是正值。随着有源区面积的减小，内建的压缩应力将增加。机械应力的增加导致晶体对称结构受到破坏。明显的形变位能变化可能会导致六重导带的破坏和两重价带的退化。载流子有效质量和能带散射率受到影响，将直接影响迁移率，导致 PMOS 器件迁移率的净增加以及 NMOS 器件迁移率的减小。SA 与 SB 的含义如图 4-4 所示。

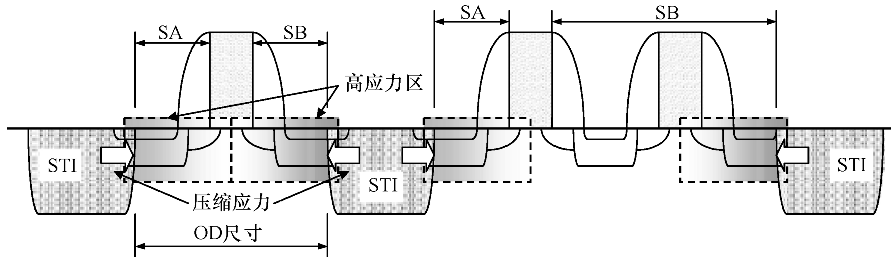


图 4-4　单个栅和多指状栅的 NMOS 晶体管的 STI 应力

STI 外径较小时，STI 应力效应将导致空穴迁移率的增加与电子迁移率的减小。图 4-5 所示为因 STI 应力效应产生不同类型的 I_{DS}变化曲线。建模是针对图中外径的交叉点。实际电流与模型电流之间的变化程度很重要，因此，模型必须包括 STI 应力效应以确保仿真结果能很好地反映硅器件特性。从这些数据显然可见，克服这一效应的唯一方式是增加源和漏区的长度，但是将栅分成多个指状也可用来增加总的 OD 尺寸，如图 4-4 所示。

在 0.15μm（也可能是 0.18μm）工艺时，栅泄漏开始成为一个必须考虑的问题。对某些特定的模拟元件，如环路滤波电容，解决问题的方法是不得不使用厚氧器件、多层布线的金属电容或者金属梳状电容来解决。在小于 0.13μm 的工

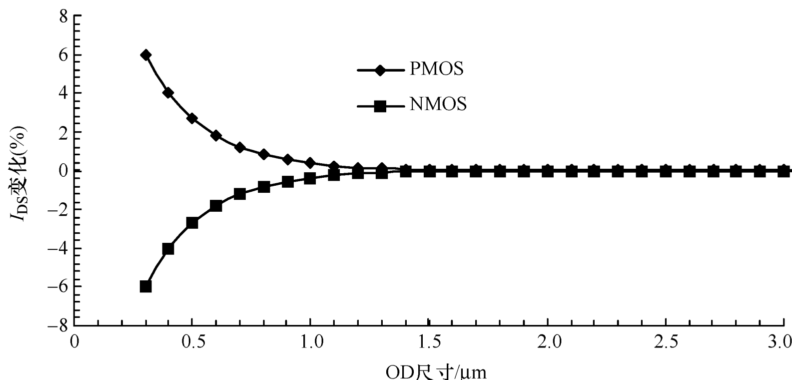


图 4-5　先进工艺（0.13μm 或更小）$I_{d\ sat}$的变化

艺中，其他电路的栅泄漏也必须考虑。90nm 工艺中，25℃时，采用 V_{cc} 偏压的尺寸为 15μm × 0.4μm 的晶体管，栅泄漏电流将高达 5nA。在许多应用中，如简单的数字逻辑中，这么小的电流并不会严重影响电路的运行，但是低功耗设计中，偏置电流非常小，器件工作在亚阈区或非常接近亚阈区，这些泄漏电流将变得非常重要，因此在设计过程中必须加以考虑。在采样-保持电路与多米诺逻辑中也必须同样考虑泄漏电流问题，因为泄漏电流泄放掉电压保持节点的存储电荷，将引起节点电压漂移。另外，栅泄漏电流源于 Fowler-Nordthein 隧穿，可以被建模为散粒噪声，这是一个可以引起白噪声的随机效应[38]。这个噪声将影响某些电路（如低噪声放大器和压控振荡器）的整体性能。栅泄漏可以分成三部分[2]：两边的直接隧穿（Edge Direct Tunneling，EDT）部分（I_{gs0}与 I_{gd0}）以及栅到沟道的隧穿（I_{gc}）（见图 4-6）。对长沟道器件，栅到沟道的隧穿占主导地位。一般情况下，边界直接隧穿小于栅到沟道的隧穿，因此，随着沟道长度增加，单位泄漏电流将增加。

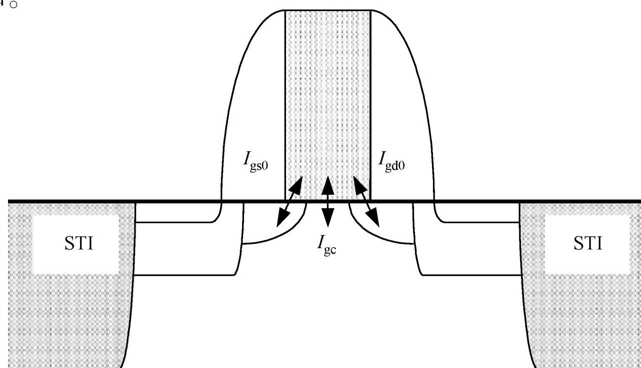


图 4-6　栅泄漏的组成

有一段时间，曾采用 pocket 注入（大斜角注入）方法减小阈值电压下降和穿通。pocket 注入的引入对模拟电路有双重影响。第一，它增加了漏感应阈值电压漂移（DITS），这在模拟和混合信号电路中经常使用的较长沟道的器件中更加显著。如果电流镜中器件的源漏电压不匹配，这个效应会在电流镜中增加更多的误差。第二个效应是使输出电阻减小，这是由于阈值电压是漏源电压的函数，并随着漏源电压的增加而不断减小。因为阈值电压对漏源电压有强依赖关系，所以必须考虑使用以下近似方程：

$$I_{DS} \approx \frac{\mu_n C_{ox}}{2} \frac{W}{L} [V_{GS} - V_t(V_{DS})]^2 (1 + \lambda V_{DS})$$

式中，μ_n电子迁移率；C_{ox}是栅氧化层电容；W 是器件宽度；L 是沟道长度；V_{GS}是器件的栅源电压；V_t是阈值电压，它是漏源电压的函数；而 λ 是器件的沟道长度调制系数。

输出阻抗的减小对模拟电路的影响非常明显，因为这将减小大部分模拟电路的有效噪声隔离。在特定情况下有必要采用“模拟晶体管”，即不使用 pocket 注入的晶体管。然而，这样做将要求增加附加的掩膜和注入步骤，因此将增加数字 CMOS 工艺的成本和生产周期。目前基于本章描述的方法仍然有可能设计出鲁棒的电路。

应该考虑的最后一个效应是 NBTI 效应对长期阈值电压漂移的影响[3]。本质上，加在 PMOS 器件上的偏压会导致界面陷阱的产生，从而引起阈值电压的漂移。不像热载流子效应，增加沟道长度并不会缓解这种效应。偏移量依赖于电路配置的细节。开关电路中偏置并不是持续加载的，因此产生的漂移较小，但是仍然会有这种效应发生。对高速时钟电路，较小的占空比变化就会导致错误动作，这仍然是一个存在的问题。阈值电压偏移可能达数十毫伏，这将带来巨大的影响。解决这个问题的一个可能的方法是在工艺角模型中包含阈值电压漂移。

4.4 无源器件

在模拟和混合信号设计中经常使用的无源元器件包括电阻、电容、变容二极管和电感等。随着工艺尺寸的逐渐缩小，电阻并没有很大的改变，只是因为掺杂浓度的改变使电阻率相对于以前的工艺有所变化。首要的关注点还是如何获得精确的模型并理解其随着工艺、温度和所加电压的变化。大多数情况下，电阻需要被建模成多个电阻，以精确描述端电阻以及寄生电容。

随着工艺尺寸的缩小，将给电容带来更多的问题，具体情况与使用的结构有关。图 4-7 所示为电路设计中使用的几种电容。在 0.18μm 工艺节点，对用作环路滤波器电容的薄栅氧器件，栅泄漏开始成为问题。25℃时薄栅氧器件中的栅泄

漏电流一般为 5nA/μm^2 或者更高。电容的典型值大约是 12fF/μm^2。如果一个 PLL 滤波器要求 200pF 的电容，总的栅泄漏电流将达到 83μA 的水平，明显大于典型的电荷泵电流。这个泄漏电流表现为 PLL 输出中过多的确定性抖动。为此不得不在环路滤波器中使用厚栅氧器件，但是由于电容将随着外加电压的变化而变化，这将会产生另外的问题。对环路滤波器来说，因为控制电压会在一定电压范围内变化，所以这个问题更加突出。为此，将会采用变容二极管或本征 NMOS 器件。变容二极管实际上是一种特殊的 MOS 电容，它采用了特殊的结构[13]。

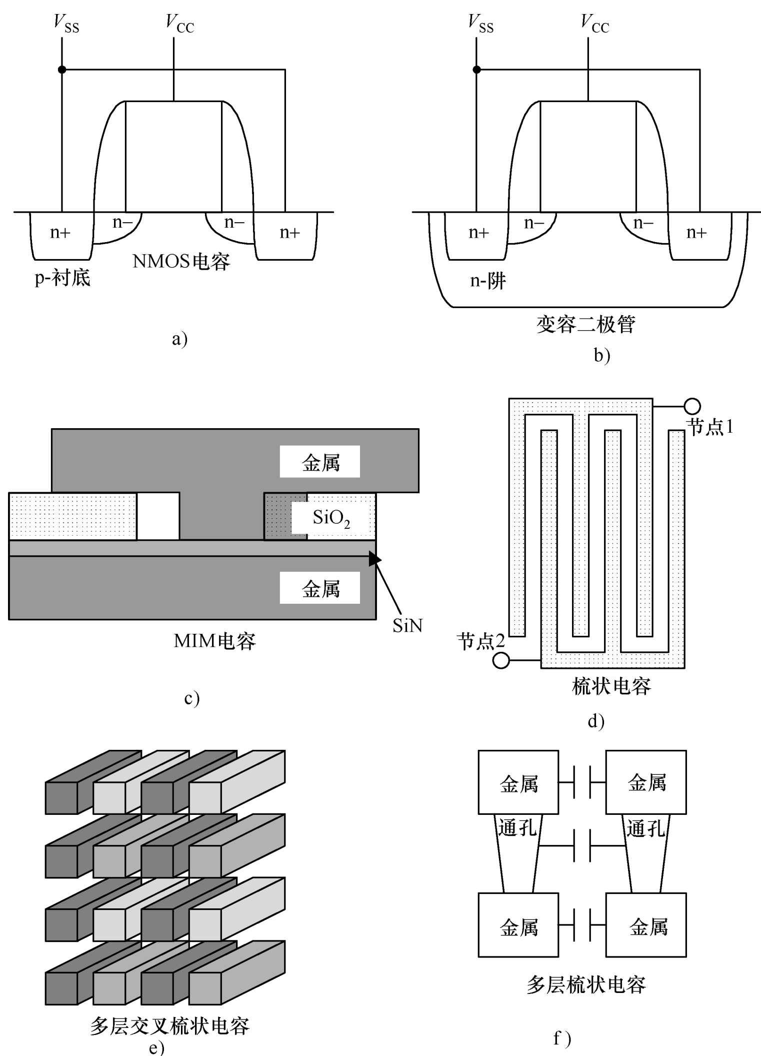


图 4-7　集成电容实例

某些情况下，如果采用混合信号工艺，也可以采用金属-绝缘层-金属（MIM）电容。这些电容一般是在两层金属之间夹一层氮化绝缘层结构，其电容约为 $1\sim4\mathrm{fF}/\mu\mathrm{m}^2$。由于 MIM 电容的寄生电阻很小，因此一个主要优点是具有较高的 Q 值（高达 150）[15]，并且其电容值是常数，与所加电压无关。

另一个常用的电容结构是梳状金属电容，这实际上是金属板－金属板电容的变形。几种不同的结构如图 4-7d～图 4-7f 所示。梳状结构是由金属形成的，因此可以在单位面积形成较大的电容，并且 Q 值较高。这些电容的容量取决于彼此相互靠近的两条金属线之间的互连电容。如果电容的绝对值很重要，那么就要使用由同一金属层两条金属线构成的梳状结构电容，如图 4-7d 所示，这种结构的单位面积电容没有多层电容那么高。图 4-7e 和图 4-7f 所示的多层金属结构，增加了单位面积电容，但是因为金属厚度变化的组合效应会增加总的电容误差。准确预测图 4-7f 所示结构的通孔到通孔电容很困难。低 k 介质的使用将减小这些结构的单位面积电容。设计金属层电容时，铜互连工艺的影响也必须考虑，因为计算实际的电容值时，必须考虑金属密度规则，并且还必须考虑作为金属密度函数的金属厚度的变化。金属密度效应很容易使金属层厚度增加 2%～5% 的误差。考虑到金属电容的各种波动因素，他们一般用在对电容绝对值的要求不是很重要的应用中。即使在这些情况下，最好在所有的工艺表征测试中能包含测试晶片，以校验实际的电容值。在参考文献［15］和［16］中还介绍了其他电容结构。

混合信号设计中普遍使用的另一种元件是电感。因为能明显改善相位噪声，在低噪声放大器和基于 LC 的 VCO 单元的集成电路中正越来越普遍地使用电感。由于电感是由金属构成的，因此还没有遇到 MOS 器件所面临的工艺缩小问题。电感的主要变化是伴随铜互连工艺的引入，通过减小串联电阻改善了电感的 Q 值。当低 k 介质工艺占主导地位时，寄生电容减小（假定这个电容并没有被用来作为 LC 电路的一部分），电感特性将会得到进一步改善。由于顶层金属厚度的增加（在标准的数字工艺中大约是 $1\mu\mathrm{m}$，而在专用的混合信号工艺中则高达 $3\mu\mathrm{m}$），因此电感一般使用顶层金属构成。由于倒装焊芯片正在成为主流，因此必须考虑与封装相关的问题。对于给定的凸点尺寸（直径为 $30\sim80\mu\mathrm{m}$），为了避免封装信号和电感之间出现不希望的耦合，有必要引入封装走线的禁止区。图 4-8 是一个螺旋电感的简化图。一个带有图案的接地面被用来减小漩涡电流损失。通过在接地面上与电感线圈直交的方向上开狭缝，由电感的磁场所感应的镜像电流将被截止，这将极大地减小负自感。图 4-8 所示集总模型中的各个元件可以作如下近似[11]：

$$L\approx\frac{37.5\mu_0 n^2 a^2}{22r-14a}\quad（螺旋电感）$$

$$R \approx \frac{l}{W\sigma\delta\left(1 - e^{\frac{-t}{\delta}}\right)} \quad（螺旋电阻）$$

$$\delta = \sqrt{\frac{2}{\omega\mu_0\sigma}} \quad（趋肤深度）$$

$$C_{\mathrm{P}} = \frac{nW^2\varepsilon_{\mathrm{OX}}}{t_{\mathrm{OX1}}} \quad（并联电容）$$

$$C_{\mathrm{OX}} = \frac{Wl\varepsilon_{\mathrm{OX}}}{t_{\mathrm{OX2}}} \quad（螺旋与衬底间的电容）$$

$$R_1 \approx \frac{2}{WlG_{\mathrm{sub}}} \quad（衬底电阻）$$

$$C_1 \approx \frac{WlC_{\mathrm{sub}}}{2} \quad（衬底电容）$$

式中，n 是匝数；r 是螺旋半径；a 是螺旋平均半径；W 是线宽度；l 是线长度；t 是线厚度；σ 是线传导率；δ 是趋肤深度，t_{OX1}是主螺旋与下面交叉点之间的氧化层厚度；t_{OX2}是螺旋与衬底之间的氧化层厚度；μ_0是真空磁导率；G_{sub}是单位面积的电导率；C_{sub}是单位面积的衬底电容。精确地计算电感必须使用三维场解决方案并且采用测试结构进行检验。

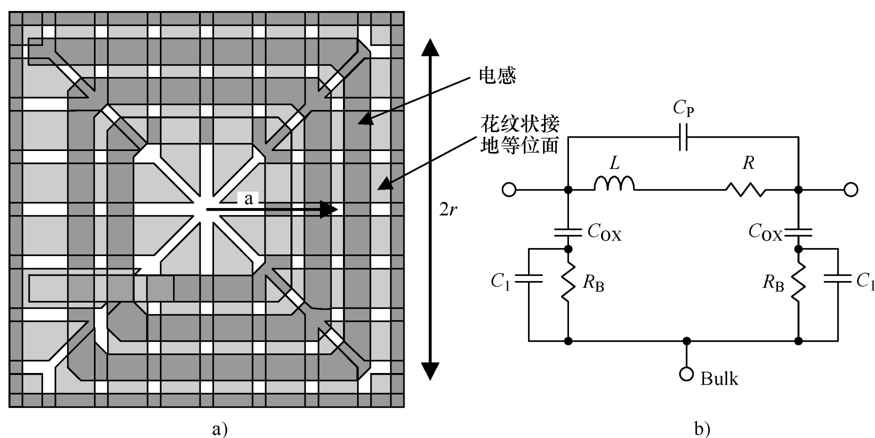


图 4-8　a）采用带图案的接地平面的螺旋电感　b）相关的集总模型

4.5　设计方法学

为了采用先进工艺成功地设计混合信号电路，要求在实施之前先进行仔细的架构研究。应该采取几个重要的步骤以确保采用这些工艺能够实现特定电路的设计。关键决策中的任何错误都可能导致重新设计，这会导致产品失去关键的上市

时间。在工艺开发阶段必须开始同时进行模拟和混合信号模块的评估，如果等到工艺决定之后才进行评估，会导致整个产品开发的滞后，也会失去改换工艺的机会。

4.5.1 工艺测试基准电路

在工艺开发阶段，非常重要的一点是确定采用新工艺实现的模拟模块，从而在开始就能够将它们加入到工艺表征晶圆中。应该通过这些模拟模块来提取关键的参数，例如，可以采用简单的运算放大器研究增益、带宽、输入工作范围、电源噪声容限、失调电压等等。也可以增加各种 VCO 单元，以查看增益、相位噪声、工作范围和电源噪声容限。还可以使用电流镜评估输出阻抗和匹配特性。另外带隙基准源也应该被考虑。关键问题是保持单元相对简单，能够用于评估工艺，而不是设计本身的细节。特别是在工艺开发初期，此时模型尚不成熟并且可能不准确，如果能够采用测试芯片实现上述评价工作就显得特别重要。分析一系列不同尺寸的器件匹配问题和配置情况以评估邻近效应也是很重要的。这些模块应该使用所有可用的氧化层厚度选项和可能的阈值电压来实现。当决定对特定的功能应该使用哪个器件时，要进行一些折衷研究。虽然测试芯片是检验设计的最好方式，但是采用一系列可以用器件模型进行仿真的电路也能分析工艺中存在的潜在问题。

4.5.2 薄氧器件设计

在设计系统的模拟和混合信号部分时，必须做出的关键决定之一是使用薄氧器件还是使用一种氧化层较厚的器件。每种方式各有优缺点。如果考虑使用薄氧器件，就要求考虑前面讨论的所有的器件问题以确保设计能成功实现。

电源裕度现在也成为一个主要问题。实际上，在这些更先进工艺中，阈值电压正接近 0.3V（假定 slow models）。假如电源电压在 1.0 ~ 1.2V 的水平，考虑到温度和电源的变化，设计者必须保持电源电压达到阈值电压的三倍以上。使用较低阈值的薄氧器件（假定这是一种工艺选项）可以得到一定的电源裕度，但是这种方式会增加仿真中工艺角的复杂度。如果使用只是阈值电压不同的两类器件，它们的模型应该是部分相关的。这两类器件的多晶特征尺寸控制和氧化层厚度应该相似，只是调整阈值电压的注入工艺不同（但是仍然局部相关）。这时理解这两种模型之间的相关情况会变得困难。因此，如果可能的话，在关键电路中应该尽量避免使用几种不同阈值的器件。

假定混合信号模块的一部分是数字电路，如 Sigma-Delta 转换器中的数字滤波电路，使用薄氧器件的优点之一是可以减小总面积。由于无需进行电平转换，该设计方案中信号到核心电路的传送更简单。对锁相环电路来说，电平转换会成

为一个特殊的问题，因为它会产生一个可能的相位误差，而且很难被正确补偿。

必须在结构上做出一些折衷，包括考虑使用薄氧器件对关键电路模块如 VCO 中的影响。对图 4-9 所示的简单 VCO 单元，两个偏压，V_{BN} 和 V_{BP}，与各自电源轨之间至少必须有一个阈值电压的差别。对初步估计，VCO 的控制电压工作范围是：

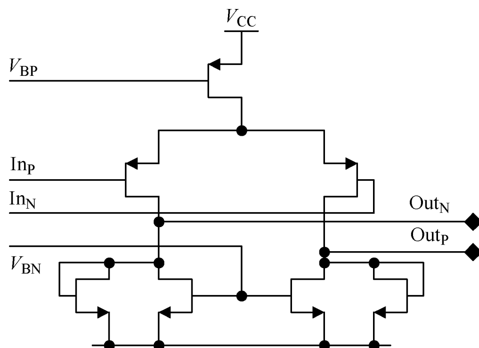


图 4-9　简单的压控振荡器单元

（来源于参考文献［12］）

$$V_{tN} < V_{control} < V_{CC} - V_{tP}$$

在 0.13μm 工艺中，V_{CC} 可能是 1.2V，V_{tP} 和 V_{tN} 可能是 0.32V，则 $V_{control}$ 的电压摆幅是 0.56V。在 90nm 工艺中，V_{CC} 可能是 1.0V，V_{tP} 和 V_{tN} 可能是 0.3V，导致 $V_{control}$ 的电压摆幅是 0.4V。如果 VCO 必须工作在相同的范围，例如 1GHz 到 300MHz，不包括任何裕量，VCO 的增益对 0.13μm 工艺是 1250MHz/V，而对 90nm 工艺是 1750MHz/V。90nm 工艺时 VCO 增益高出很多。如果不采取任何措施去减小系统的噪声，90nm 工艺锁相环的整体性能会退化，其原因是控制电压或电源轨的任何噪声都会被较高的 VCO 增益放大。

4.5.3　厚氧器件设计

第二种选择是在模拟模块和一些可能的数字模块中使用氧化层较厚的器件，然后将电平转换到较低的核心电压电平值。这种方式的优点是能增加模拟电路的裕度，可以采用电流镜和增益级的级联连接以增加抗噪声能力。栅泄漏问题会急剧减小，源漏泄漏电流也同样会减小。但是采取这种方法会带来两个问题。首先，模拟或混合信号模块可能在某点与核心部分连接，这就要求采用一个电平转换器。对数据转换器，这些电平转换器本身不会存在问题，但是需要从整个系统的角度考虑问题。数据转换器的参考时钟可能来自于另一个时钟信号。如果这个时钟来自于核心部分的电源，则必须进行电平转换。该电平转换又会带来一个额外的噪声点，增加了参考时钟的抖动或使占空比退化，这会减小数据转换器的精确度，尤其是当占空比对设计非常重要时。如前面已经讨论过的，相似的问题也会在 PLL 电路中出现。图 4-10 所示为一个可能的 PLL 结构图，其中 PLL 使用了厚氧器件并且电平转换（LB 模块）将输出时钟转换到核心电路的电压电平。图 4-11 所示为第二种 PLL 架构，其中仅核心模拟单元采用厚氧器件，当这些核心单元和 PLL 中其他数字部分之间转换时，使用电平转换。这个架构仍然有问题，因为电平转换会引起占空比问题。图 4-12 所示为一个在许多应用中采用的传统

的电平转换器，图 4-13 所示为一个最近开发的先进电平转换器，它采用厚氧本征 NMOS 器件从而避免了电压较高的 V_{CC} 电源对薄氧器件的影响。这个厚氧本征 NMOS 器件有一个非常低的阈值电压，一般接近于 0，使得薄氧器件可以作为两个电源域之间的转换点，这有助于减小厚氧器件的阈值电压问题。

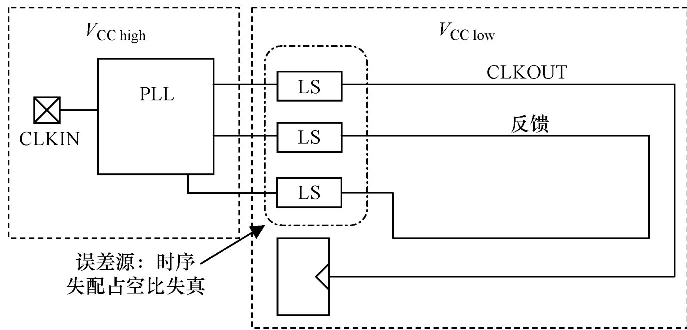


图 4-10 可能的 PLL 架构

（整个 PLL 由较高电压的电源供电而核心部分由较低电压的电源供电）

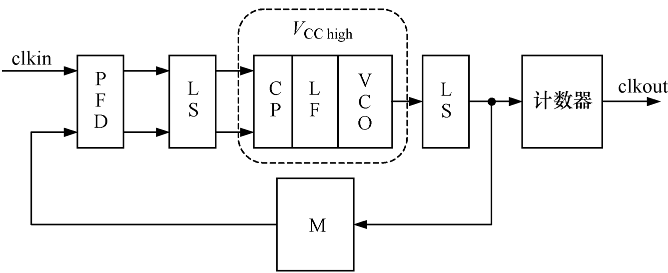


图 4-11 可能的 PLL 架构

（仅 PLL 核心部分由较高电源电压供电而其余逻辑部分由较低电源电压供电）

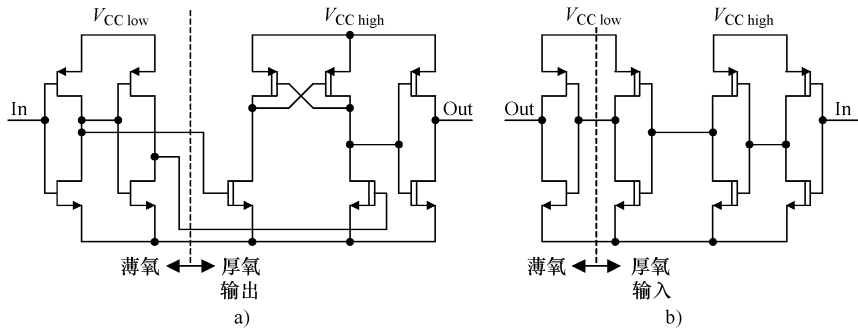


图 4-12 传统的电平转换器

a）电平抬高 b）电平降低

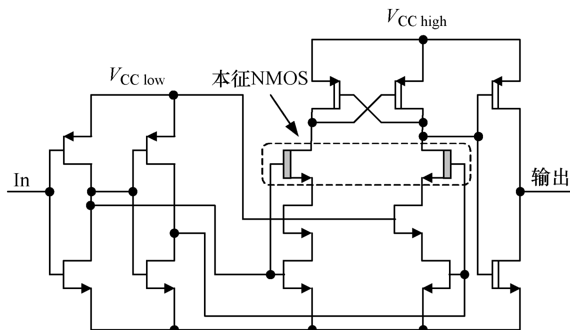


图 4-13 新的电平转换器，能对 1.2V 器件提供较大的保护
（来源于参考文献［17］）

从全局设计的观点来说，最好是所有电路都与核心电压匹配，特别是如果可能的话，锁相环也应该这样。这一目标的实现一方面取决于特定的应用情况，还要考虑应该如何调整架构以缓解每一种方式存在的问题。

4.6 低压技术

在过去的几年，设计低工作电源电压的模拟和混合信号电路的许多方法已经得到了改进。由于氧化层厚度更薄，要求采用更低的电源电压。例如便携式产品，如手机，都使用较低电压的电源。这更增加了低工作电源电压的模拟和混合信号电路设计方法的需求。虽然电源电压降低了，但是器件阈值电压并没有等量降低。在早期的工艺中，因为级联连接有较高的输出阻抗，被用来形成高增益电路。随着电源电压的减小，这种结构变得难以实现，因此设计者被迫采用多级相连的方式以获得同样的增益。虽然这一技术起了作用，但是每个增益级都增加了附加的极点和零点，使补偿变得更加困难[5]。下一节将介绍几种用于模拟和混合信号设计的不同低电压组成模块。

4.6.1 电流镜

电流镜是模拟和混合信号电路中最基本的组成模块。一个好的电流镜具有优良的对参考电流的复制能力以及能够抗噪声的高输出阻抗。过去这是用级联连接实现的，但是这可能无法考虑减少裕度。为此又开发出其他的电流镜，在维持高输出阻抗的同时可以减小裕度问题。这些先进工艺中的晶体管输出阻抗已经衰减的相当多了。图 4-14 所示的电路就是一个低压、高输出阻抗电流镜的实例。通过设计器件的尺寸，例如使 MN2 的尺寸为其他器件尺寸的四分之一，能够将所需的输出电压减小到 $2V_{d\ sat}$。这个电路有几个问题。第一，由于

MN1 与 MN5 的 V_{DS}值不相等导致电流匹配不完美。在小尺寸工艺中由于阈值电压对源漏电压敏感性的不同使这个问题被进一步放大。因为它们会导致运算放大器偏置电压的增加，所以必须考虑这一问题。另外，由于源漏电压不同，各种器件在工作时间内阈值电压漂移会不同，将导致失配进一步增加。

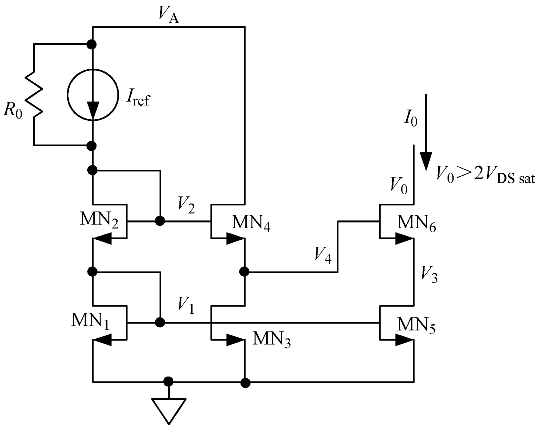


图 4-14　低压高阻抗电流镜（来源于参考文献［4］）

图 4-15 所示为另外两种改善了电流匹配能力的低压电流镜。图 4-15a 所示的电流镜已经广泛使用好几年了。通过控制 V_{CAS}，可以在实现维持高输出阻抗和良好电流匹配的同时使输出电压摆幅最大。图 4-15b 所示的电流镜具有相似的性能而且输出节点维持在固定电压以减小电路参考一侧的电流误差[6]。通过控制相应的参考电压，可以实现输出节点在较宽的范围内摆动。在使用图 4-15b 所示的电流镜时必须倍加注意有源输入要求稳定的补偿。

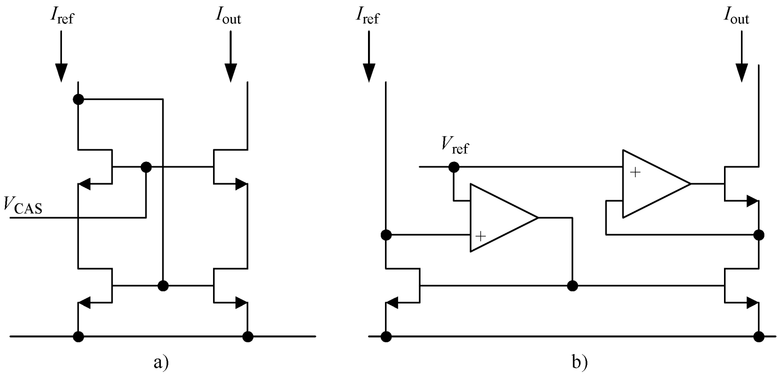


图 4-15　改善了电流匹配的附加低压/电流参考的电流镜

图中所示的电流镜在减小裕度的同时甚至还提供了使用级联结构的能力，但是栅泄漏效应一直被忽略了。在实际实现设计时必须考虑栅泄漏问题，因为它会影响电路工作，特别是工作在弱反型时，这时参考栅泄漏电流较小但是器件栅泄漏电流较大。

4.6.2　输入级

目前已出现了几种新颖的低压轨到轨输入级结构[8,9,10]，它们普遍采用了两种方式。第一种方式如图 4-16 所示，采用 PMOS 和 NMOS 差分对。对接近于 V_{SS} 的低输入电压，仅 PMOS 差分对被偏置到开启状态。对接近于 V_{CC} 的很高的输入电压，仅 NMOS 差分对被偏置到开启状态。这会导致作为共模函数的增益的变化，而这是不希望的。

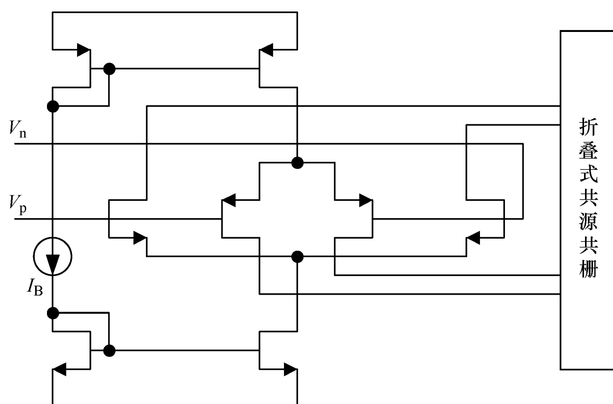


图 4-16　使用 PMOS 和 NMOS 差分对的轨到轨输入级

另一种更新颖的输入级是体驱动 MOS 晶体管。图 4-17 显示的是使用这种技术的简单电路。通过驱动器件的衬底，可以允许在输入级输入负电压、零电压或较小的正电压，这样能增加输入共模范围。然而这一结构也存在几个缺点，包括：与传统的差分对相比跨导的急剧减小；最大工作频率的减小；可用于输入对的器件可能受到限制（例如，除非使用深 n 阱工艺，否则不能使用 PMOS 器件）；晶体管必须被放置在一个独立的 n 阱中（假定使用 PMOS 器件），这会引起失调电压增加；开启寄生晶体管的闩锁问题；和较大的等效噪声等。

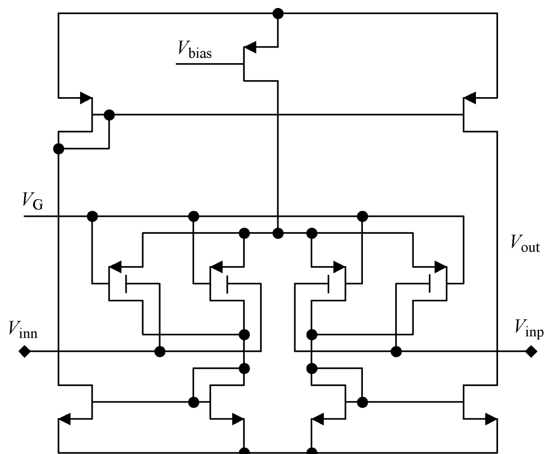


图 4-17　运算跨导放大器的衬底驱动输入级

4.6.3 输出级

输出级也必须提供轨到轨的工作能力，特别是对低压应用。图 4-18 所示为三种不同的适合低压轨到轨工作的 AB 类输出级。图 4-18a 所示为一个非常简单的输出级[10]。静态电流由电压 V_Q设定。图 4-18b 所示为 Monticelli 输出级[36]，它在低压应用中被广泛使用。输出级能进一步提高放大器的增益，这是我们所乐见的。最小电源电压是 $2V_{th}+3V_{DS\ sat}$，这与图 4-18a 中所示的输出级相同。最后，图 4-18c 是 You 等人描述的输出级。虽然这个输出级的增益没有 Monticelli 输出级那么大，但是它能在低到 $V_{th}+2V_{DS\ sat}$的电源电压下工作。

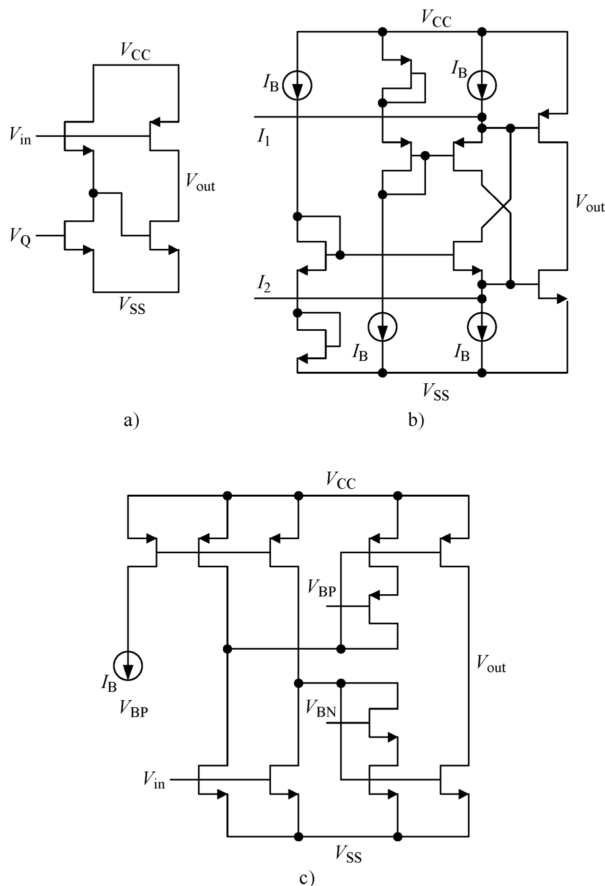


图 4-18 各种低压应用的轨到轨输出级

4.6.4 带隙基准

随着电源裕度的持续减小，带隙基准设计就变得越来越困难。图 4-19 所示电路是针对这一问题提出的一个“聪明”的解决方案。将两个偏置节点分压，再将

运算放大器连接到分压节点上。由于是运算放大器，节点 N1 和 N2 的电压相等。如果使电阻 R_{2A1} 和 R_{2A2} 分别与 R_{2B1} 和 R_{2B2} 相等，则节点 N3 和 N4 的电压也相等。

如果运算放大器中采用 PMOS 构成的差分对，由运算放大器限制电源裕度，则允许的最小电源电压可以表示为

$$V_{\sup\min}=\frac{R_{2B2}}{R_{2B1}+R_{2B2}}V_{EB2}+\left|V_{thp}\right|+2\left|V_{DS\,sat}\right|$$

或者如果是 pnp 晶体管和电流源限制电源裕度，则：

$$V_{\sup\min}=V_{EB2}+\left|V_{DS\,sat}\right|$$

再结合前面结论，可以得到下式：

$$V_{out}=\frac{R_3}{R_2}\left(V_{EB2}+V_T\frac{R_{2A1}+R_{2A2}}{R_1}\ln M\right)$$

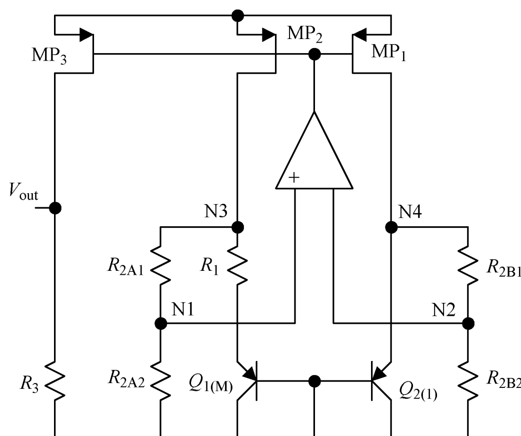


图 4-19　低压带隙基准

（来源于参考文献［7］）

因为 MP_3 漏源电压与 MP_1 和 MP_2 的不匹配，这个电路也还存在一些问题。采用几种方式的套叠可以减小上述失配问题，但是电源限制会阻碍这种连接方式的实现。对非常先进的工艺，二极管路径已证明是一个限制因素，特别是低温时。

4.7　设计过程

伴随着工艺的先进，集成电路设计的一个主要问题是要考虑版图设计前后的差异。当各个节点的总负载中互连寄生占主导地位时，这个问题就出现了。最初，互连效应被完全忽略（0.5μm 和更大尺寸的工艺）。对较大的几何尺寸，虽

然忽略了互连效应，但是电路仍然能正常运行，因为器件模型已包括了寄生效应。当几何尺寸持续缩小时，互连在电路节点总负载中所占的比例更大。在 0.25μm 工艺节点附近，设计者必须考虑互连的寄生效应以确保电路功能正常。不过，除了很长的信号线之外，互连的阻抗一般可以忽略。而在 0.15μm 工艺，即使对更局部的互连，寄生电阻也很重要而必须考虑。对 90nm 或更小的尺寸，必须提取全部寄生参数网表以确保电路功能，但是这要求在参数提取时必须保证电阻电容提取（RCE）的精确性。在处理版图设计前期、后期的问题时，可以采用下述 3 项基本技术：

（1）尝试在版图设计前期对所有问题进行建模，以缩小版图设计前期、后期的差异。

（2）不建立特殊模型；尽可能快的完成版图初步的设计，然后再对该版图进行优化

（3）使用历史数据进行版图设计前期的评价，以尽量满足设计目标。然后尽可能快地进行版图设计，再对设计的版图进行优化。

如果采用上述方法（1），版图设计前期的建模最耗费时间，并最容易出错，因为在设计的早期阶段就确定版图将是什么样子，即使不是不可能的话，有时也是很困难的。如果没有实际的版图，甚至没有一个布局规划，评估布线的寄生参数会变得很困难，除非这个单元很简单。采用方法（2）能够以最快的方式开始版图设计过程，但是最终的布局布线与开始进行的布局布线之间必然存在较大差别（对一个像压控振荡器这样的单元，产生的差别可能约为 35% ~45%）。这里最关键的问题是产生一个基本的设计，进而产生一个能供以后进行模拟的版图，但是在一个设计定案之前，需要反复进行提取版图网表的过程。其中困难之处在于处理版图产生的网表，而这种网表的可读性往往较差。因此，需要采用合适的方式建立电路层次，即主要是对多个较小的模块提取网表，从而使网表更易于管理。方法（3）其实就是方法（2）。只是在方法（3）中，前一代工艺中采用单元信息被作为一种通用的准则，以确定需要多少余量设计使设计满足规范要求。如果在 0.13μm 工艺中，已知一个给定的单元，如高速计数器，在版图设计前后差别为 40%，这可以作为下一代工艺中关于可能发生变化的一般性准则。这种方式迅速，并且其精确度并不比花费大量时间试图在版图设计前对所有问题精确建模得到的结果差。

针对 STI 应力效应，可以采用几种处理方法。一种方法是把重点放在器件尺寸的调整，以保证与用来建立 SPICE 模型的器件版图相匹配。使用这种方法可以使应力效应的前仿真与后仿真之间的差异最小。第二种（可能更好的）方法是尝试在实际的版图之前准备一个初始的布局规划，从而在前仿真中就能包括这些效应，在实际版图完成之前获得 OD 参数。通过版图之前的布局规划，OD 效

应实际上可以用来增加如锁相环中的 VCO 或时钟恢复电路等关键电路的速度。至少它可以在版图之前考虑使用更精确的模型。因为版图对 STI 应力效应的影响很大，在版图开始之前必须对布局规划进行仔细的考虑。幸运的是，许多能够较好产生模拟版图的技术也能够自动产生与 STI 应力效应的良好匹配。图 4-20 所示为一个简单的电流镜及其两个可能的版图设计结果。图 4-20b 所示的版图在尺寸上更紧凑，但是从光刻的观点和 STI 应力效应的观点来看，它的匹配性较差。图 4-20c 所示版图采用哑元器件使 STI 应力效应最小化并且可以改善光刻和刻蚀问题，使匹配有了改善，但是占用的芯片面积较大。两个版图中并没有显示在器件的顶部和底部应该如何设计。一般在临近的单元之间增加保护环以保持单元两侧版图的连续性。

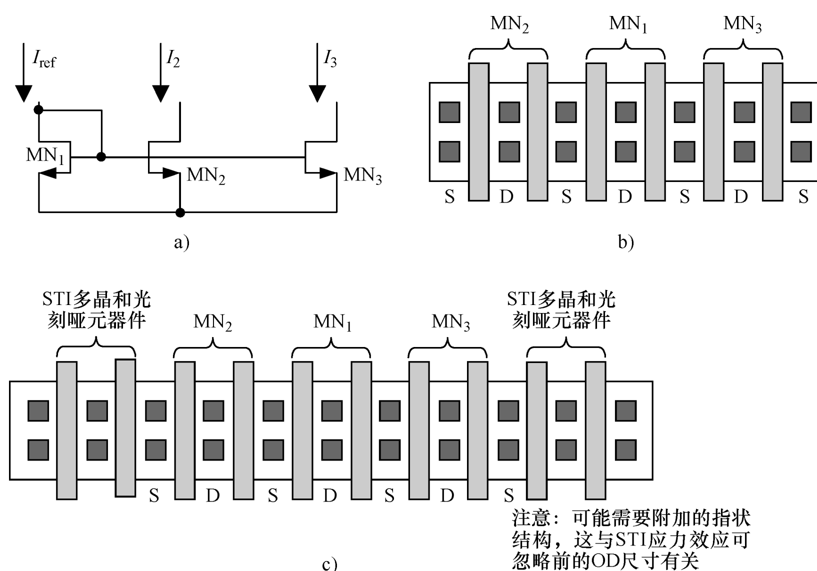


图 4-20　a）简单的电流镜电路　b）芯片面积小，但是匹配差　c）采用哑元器件改善版图设计

4.8　静电放电保护

在更先进的工艺中，由于栅氧化层越来越薄，关键模拟和混合信号模块的静电放电（ESD）保护变得日益困难，这一问题与模拟模块具有比数字模块较轻的负载有关。器件的击穿电压正日益减小，对 130nm 工艺节点已经达到 6V 之下。ESD 保护必须在期望的触发点能快速导通，并将电压保持在期望的水平。考虑到模拟和混合信号电路正不得不与更大的数字电路集成在一起，这就要求对模拟和混合信号部分开发 ESD 保护方案时必须考虑整个芯片的拓扑结构。

4.8.1　针对多电源情况的考虑

图 4-21 所示为一个针对芯片中使用多个地电平情况的保护方案实例。因为 ESD 作用可能在任意两个引脚之间发生，可以用背靠背二极管将各种接地连接在一起，给 ESD 泄放电流提供一条泄放路径。采用两个二极管可以减小各种接地之间的噪声耦合，在决定二极管尺寸和串联的二极管数目时，必须考虑这种耦合。必须确保在电路正常工作期间二极管从不正向导通，否则会使模拟电路的性能严重退化。图中作为实例，显示了几个 ESD 放电通路，表示 ESD 作用下不同元件的采用情况。

图 4-21　基于较大数字 IC 构成的 PLL 的 ESD 保护配置
[该数字 IC 中 I/O（VSSN）、核心电路（VSSQ）、PLL 模拟（VSSA）和 PLL 数字（VSSD）等分别采用不同的接地]

为了适应先进的高性能微处理器的时钟频率，I/O 信号频率不断增加。这就要求减小压焊点的寄生电容，但是对 ESD 保护器件的要求并没有免除。随着栅氧化层的减薄将导致 ESD 电容的减小，但是决不能因此而影响 ESD 保护水平。实际上，保护级别必须保持不变甚至更好。这对之前就已考虑这个问题的射频（RF）电路设计者来说已经不是什么新课题了，但是在具有高速外部 I/O 的混合和微处理器设计中，我们正经历同样的困难。关键问题是输入和输出引脚采用的

ESD 保护都不应该能影响芯片基本功能的所有性能。这意味着 ESD 结构的电容必须保持不变，与工作电压无关。如果不能实现这点，就会引起谐波失真，这将可能影响高速数据转换器、RF 合成器、时钟数据恢复电路、或锁相环的性能。

图 4-22 显示了用于保护高速和模拟输入的两种不同结构。因为二极管直接连接在高速引脚，可以使节点电容保持较小。虽然器件电容可以被保持到较小，但是因为金属互连和封装寄生，整个 ESD 电容会变得较大，这在任何时候都会给总节点电容增加 200fF 到 1pF 的电容。

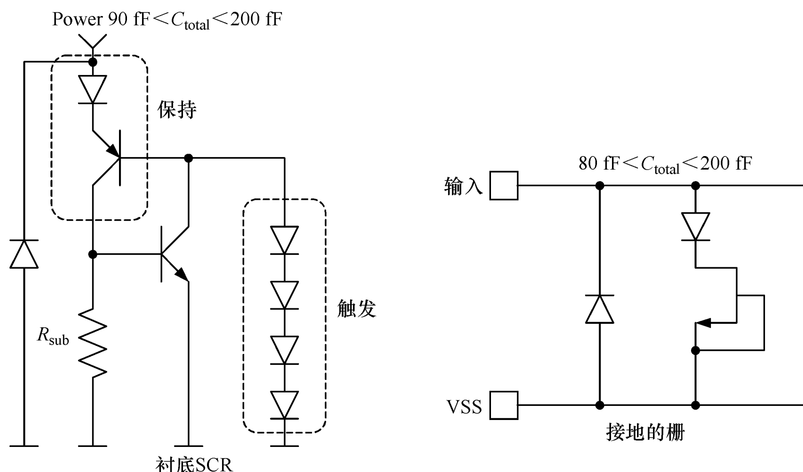


图 4-22　可以用来保护高速输入的各种 ESD 结构

4.9　噪声隔离

随着数字逻辑密度的日益增加和更快的信号边缘速率，将导致瞬态电流的实质性增加，引起更大的噪声水平。因此，对较大的数字集成电路，模拟和混合信号的噪声隔离变得更加困难。随着电源电压减小，噪声裕量更小，使这一问题更加严重。

4.9.1　保护环结构

保护环被广泛用来减小衬底中的噪声。已经进行了很多研究来探寻各种配置的优点[18-35]。关于如何更好的设计保护环结构使系统不同部分之间的耦合最小化方面，经常会有相互矛盾的信息。目前，混合信号芯片已经成为应用最广泛的产品，而不是只在个别场合才有所应用。即使现在的微处理器也包括了一些如 PLL 这样的模拟电路。接收器和发射器芯片可能包括大量的模拟和数字模块。而

芯片中模拟和数字部分之间的耦合会产生不希望的噪声，导致性能退化。

作为一个例子，考虑一个简单的模-数转换器电路。如果有用的输入范围是 2V，并且我们使用的是 12 位转换器，则一位的分辨率为 488μV。如果通过衬底来自电路数字部分的耦合噪声足够大，就会减小转换器的位有效分辨率。目前衬底耦合的噪声已经十分高了，在一些情况下达到毫伏的范围。为了设计高性能的集成电路，必须在设计阶段就考虑如何隔离衬底噪声。

一般来说，较宽的保护环将提供更好的隔离[31]。这是因为增加了噪声源与受影响的敏感电路之间的阻抗。噪声源与保护环/衬底引出端越近，噪声越低[31]，这是因为在构成噪声的杂散载流子有机会通过衬底扩散之前，就已经在靠近源的附近被轻易地收集了。噪声源与接收器之间的距离也有着十分重要的影响[31]。根据参考文献［23］的结论，隔离效果与分开的间距成线性关系，从阻抗的角度来看这是非常有道理的，因为电流的主要部分靠近表面流动，这很容易被保护环有效的收集。最终的性能和必需的间隔取决于衬底掺杂。最优的间距必须根据工艺信息来决定。对隔离电路有几种不同的策略。采用外延结构时，对被噪声困扰的电路，衬底上的保护环和紧临的隔离作用不大。对非外延结构，可以采用 P + 保护环、P + 保护环与 n 阱环、二氧化硅沟槽、SOI 和深 n 阱结构。这些保护结构的实际效果好坏取决于电源总线方案的设计与整个工艺规范（掺杂等）。许多研究表明，添加 n 阱保护环对改善隔离的作用很小，并且实际上会使噪声性能退化[19,31]。

对轻掺杂衬底，采用背面连接实际上会增加噪声[24]。这可能是由于背面连接引起阻抗增加，从而使载流子更可能流向敏感的电路而不是流出器件。抗噪声能力极大依赖于敏感电路相对于噪声源的位置、晶圆厚度和背面连接的压焊配置。而对于外延片，由于衬底的阻抗较低，情况将会不同。

与保护环的连接也是很重要的因素[28]。减小保护环连接的感抗能够极大地改善隔离效果。单独采用一个保护环引出端将有助于进一步改善隔离效果[23]。在低速电路中，衬底建模为阻性网络就足够了，但是在非常高速的 IC 中，衬底并不能被建模为纯阻性网络[18]。对高速电路采用 Pisces 和 Medici 进行仿真会因为较长的仿真时间而变得并不可行，但是这些工具可用于洞察通过衬底的耦合。

使用保护环的前提是衬底中能提供一种收集杂散载流子的方式，使它们不会到达电路的敏感部分引入不希望的噪声。设计保护环结构的关键因素之一是理解设计的对象和目标以及正确建模的能力。在许多实例中，因为对物理设计采用了不合适的模型可能导致不正确的结论。图 4-23 所示为一个可能的耦合模型的简化电路图。左边部分电路表示的是一个分别采用 n 阱和 p + 保护环与外部电源相连的噪声发生器电路。如果保护环路径上的电阻大于穿过衬底到达噪声敏感电路的路径上的电阻，噪声将耦合到敏感电路。同样，如果所有的保护环连在一起，

电流可能会从其中一个保护环流出，通过另一个保护流进与该保护环相连的敏感电路。如果封装和芯片寄生参数中包括感抗，这种可能性将大大增加，因为感抗会极大的增加高频开关噪声的串联电阻。

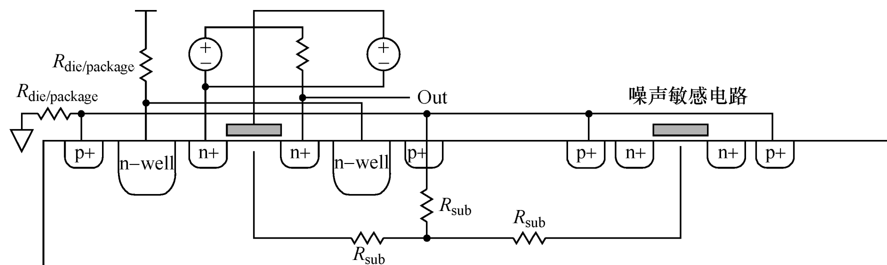


图 4-23　衬底耦合模型

4.9.2　隔离的 NMOS 器件

目前一个较新的进展，是采用深 n 阱给 NMOS 器件提供更好的隔离。该结构如图 4-24 所示。深 n 阱区是采用离子注入形成的，然后在形成 n 阱时再形成隔离的 p 阱。如果杂质浓度设计合适，位于隔离的 p 阱中的 NMOS 器件将会具有与标准 NMOS 器件相同的特性。关键问题是确保隔离区中器件的性能不产生退化。这个结构对 NMOS 器件提供了优良的隔离。频率低于 500MHz 时，达到 20 ~ 25dB。而频率高于 1GHz 时，为 5 ~ 10dB。设计使用深 n 阱工艺时应加倍小心，以避免使深 n 阱过大，否则将增加衬底的寄生电容，导致不希望的噪声被耦合到关键电路。

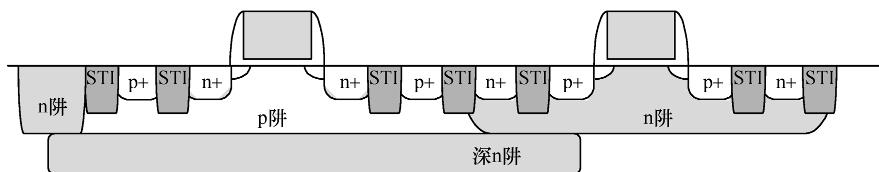


图 4-24　在采用深 n 阱工艺中的 NMOS 和 PMOS 器件的简单剖面图

4.9.3　外延材料与体硅

外延晶圆提供了低阻衬底，非常有助于减小闩锁的可能性，但是也会有效的使所有器件短路。因此，噪声较大的数字模块将与敏感的模拟模块耦合（除非采用三阱工艺）。由于这个耦合，芯片上所有的模块最好采用单一的公共地，因为它们已经通过衬底被有效地短接在一起了。设计关键模块时必须具有这种对接地端重耦合知识的了解，以确保设计的电路对接地端噪声的敏感性最小。一般来

说，对于未采用三阱工艺的外延工艺，混合信号设计是困难的，特别是如果芯片的数字部分又较大时。

4.10　去耦

对深亚微米工艺来说，因为器件密度更大、边沿速率更快、电源电压更低，去耦已经越来越成为一个问题。对高速输出缓冲供电，瞬时 *IR* 压降已经达到伏特的级别。而对于低电压的内核电源，瞬时 *IR* 压降则达到了几百毫伏。对给定的电路模块，*IR* 压降源于电源轨上的串联电阻和感抗。图 4-25 所示为一个有 IC 的电路板简化图。在板级增加的去耦元件可以减小板级的电源纹波，但是由于器件封装和芯片寄生参数的影响，在芯片级增加去耦所起作用一般极小。各种寄生参数的确切值很难提供，但是对引线键合封装，电感的范围是 5～10nH。对倒装焊芯片封装来说，电感小于 1nH。图中显示的模拟与数字部分的电源被完全隔离。

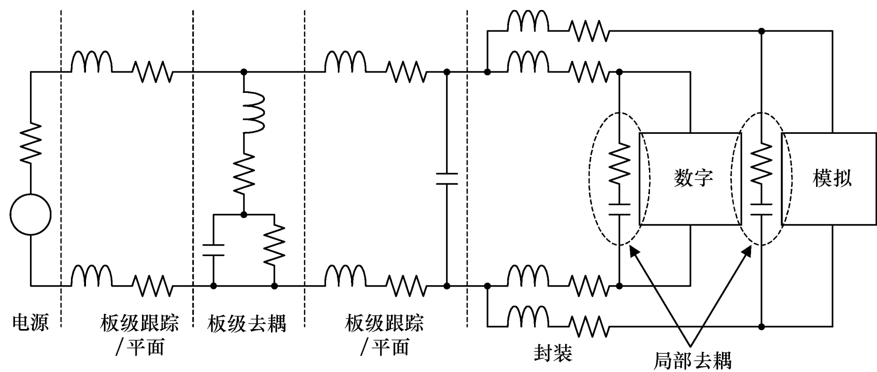


图 4-25　带有 IC 的电路板

采用下述两个简单的方程可以分析电感和去耦对电压降的作用：

$$i = C\frac{\mathrm{d}V}{\mathrm{d}t}$$

$$V = L\frac{\mathrm{d}i}{\mathrm{d}t}$$

考虑串联电感为 1nH、在 30ps 中电流变化为 1mA（对 0.13μm 和小于 0.13μm 工艺来说，这并非是个不现实的数字）的简单情况，电感上的 *IR* 压降约为 33mV，这个值的影响已经很重要了，特别是考虑电流会变到多高时。

图 4-26 所示为一个 NMOS 去耦电容的简化图。因为与其他去耦方式相比，单位面积栅氧电容的容值较大，由此，栅氧电容是最普遍用作去耦单元的器件。设计去耦单元时必须考虑寄生参数，如果忽略它们将导致不正确的假设，即认为

插入单个大面积的栅就可以形成较大的去耦单元。实际上，这样形成的去耦单元较差，因为栅长两端沟道电阻将会增加，在电容上引起较大的串联电阻。另外，多晶硅密度规则也不允许这种情况。实际上，仅在靠近边缘的栅区对电容有贡献。使这个效应最小化的唯一方式就是使沟道长度更小。这必须与使用的单元数和每单元的电容相匹配。总的沟道电阻可以从 V_{DS} 较小区域范围的 $I_{DS}-V_{DS}$ 曲线上提取。

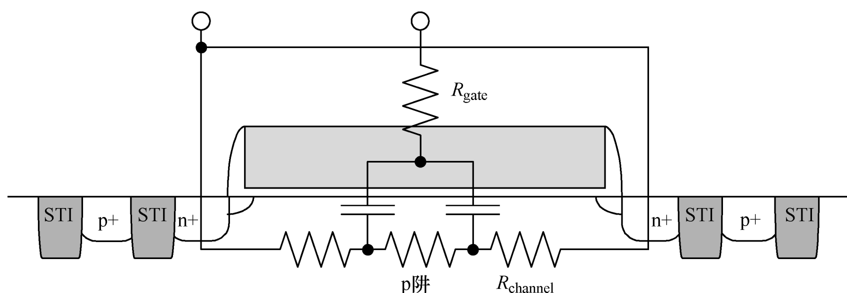


图 4-26　NMOS 栅电容

图 4-27 所示为去耦单元的简化模型。通常采用多个去耦单元并联，等效于图 4-27 右侧所示的电路。将多个去耦单元并联，就可能获得较小的有效串联电阻和较大的电容。最右侧的电路被用来模拟几种测试情况以说明封装寄生和去耦效应。假设电流源产生一个上升沿和下降沿为 30ps、峰值为 10mA 的电流（这在 90nm 工艺种是有可能的）。图 4-28 所示为电流源的电流变化时电源电压上的感抗效应。图 4-29 所示为当去耦单元的串联电阻较高时去耦是如何的无效，这说明了使电容单元的寄生电阻最小化或确定并联单元足够多的重要性。图 4-30 说明了使串联电阻最小化是如何允许电容给电路提供必需的电流，减小电源压降的。表 4-2 总结了这些测试情况下的数值。

表 4-2　去耦仿真结果简要总结

	测试条件									
	1	2	3	4	5	6	7	8	9	10
L/nH	0	0.1	1	0.1	1	1	1	1	1	1
R/Ω	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
C/pF	0	0	0	10p	10p	10p	20p	40p	20p	40p
R_C/Ω	0	0	0	1	1	10	1	1	10	10
Delta	0.01	0.143	1.343	0.050	0.077	0.130	0.040	0.024	0.110	0.106
Max. V_{CC}	1.20	1.267	1.867	1.220	1.234	1.220	1.216	1.207	1.210	1.210
Min. V_{CC}	1.19	1.123	0.523	1.170	1.157	1.090	1.176	1.183	1.101	1.105

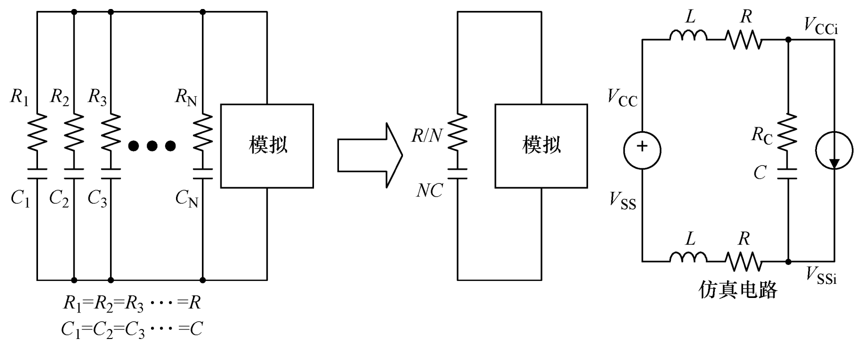


图 4-27　等效电容的定义

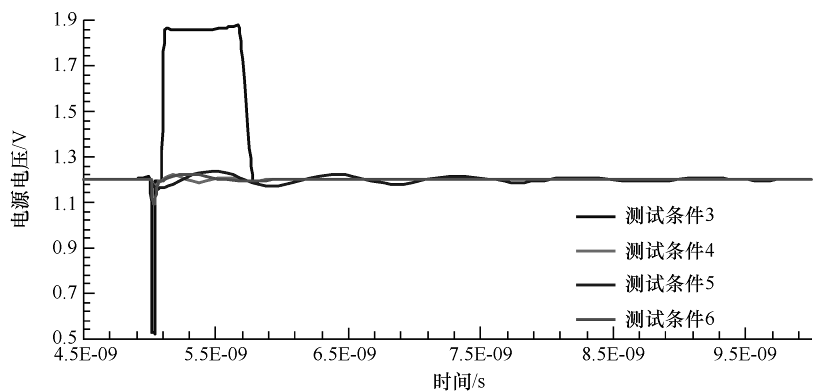


图 4-28　电源电压在无去耦和高感抗（测试条件 3）、低感抗和有一定去耦（测试条件 4）、高感抗和有一定去耦（测试条件 5）与高感抗和有一定去耦以及高串联电阻（测试条件 6）等条件下的变化

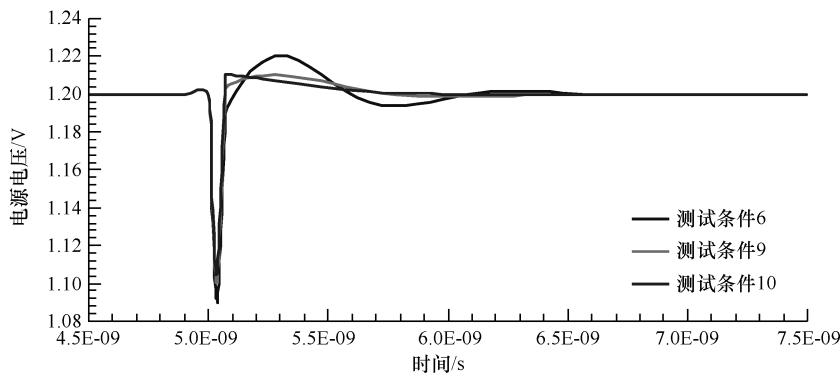


图 4-29　电源电压在高感抗、高串联电阻和去耦电容分别为 10pF（测试条件 6）、20pF（测试条件 9）和 40pF（测试条件 10）等条件下的变化

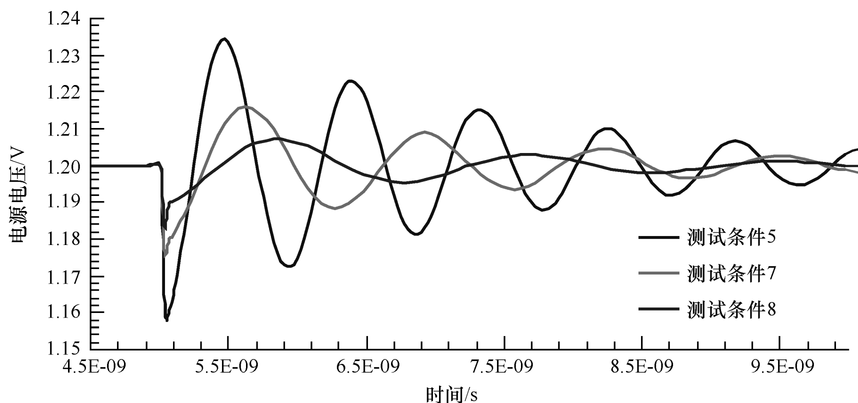


图 4-30　电源电压在高感抗、低串联电阻和去耦电容分别为 10pF（测试条件 5）、20pF（测试条件 7）和 40pF（测试条件 8）等条件下的变化

可以采用多种方法在设计中增加去耦。主要的两个方法如图 4-31 所示。去耦填充方法是在所有的空白区域填充去耦单元。只有在完成版图设计后才添加去耦电容的情况下才能采用这种方法。但是对某些电路，如 VCO 单元，一般需要用一个较宽的去耦电容环围绕其模块。第二种方法是将去耦电容分散到各个单元。这种方式要求在版图设计前进行更多的前端布局规划和仿真，但是许多情况下会获得比第一种方法更好的性能，因为去耦单元被贴近放置在需要去耦的单元旁边。确定去耦单元的数量有一点困难。一种方式是首先提取模块的平均功耗，然后使用 $C_{decoupling}=10P_{avg}/V_{CC}^2 f$ 来评估所需去耦的数量。这里必须使用最大的工作频率。这给确定去耦数目提供了较好的出发点。但是，在实际上，用于评估电源总线寄生所进行的仿真是确定是否已提供了充分去耦的最佳方式。

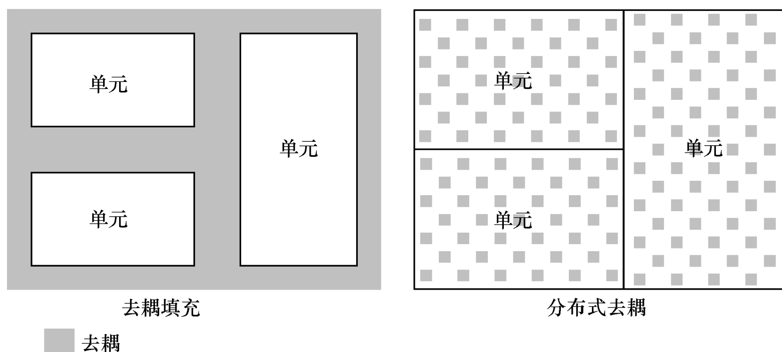


图 4-31　两种去耦方案

4.11 主电源线

对模拟单元，电迁移会成为问题，特别是给定了局部区域的金属尺寸设计规则时。由于数字电路中电流是双向的，只有自加热是需要考虑的主要问题，因此对大多数数字逻辑而言，电迁移在信号线上不会成为问题，仅需在地线上加以考虑。而自加热效应要求比直流电流远大的电流（均方根），在模拟电路中更典型，对数字电路一般不构成问题，除非使用了较大的缓冲器，例如时钟网络驱动器。重要的是要确认哪些部分的布线代表了电迁移的限制因素。这可能是接触孔、薄金属或通孔。大部分工艺技术确定金属宽度是电迁移限制因素，但是这应该得到确认，以确保如果通孔是限制因素就无需采用更宽的金属来满足针对电迁移的设计规则。

模拟单元中连续的直流电流一般比数字电路中的开关电流更成问题，虽然也必须关注主电源线以确保它们能提供必需的开关电流。在设计开始阶段必须确定电流大小以确定版图设计规范。这些设计规范必须包括接触孔最少数目、每层金属的电源线最窄宽度和两层之间通孔的最小数目。应采用产生最大电流的工艺拐点来确保已提供了充分的信号走线。一般在计算值上再增加 20%，但是余量的大小应该考虑电迁移规则的严格程度。

确定必需的主电源线的一个简单方式是在最坏电流拐点（一般对应于低阈值电压、高电源电压、低温度情况）对模块进行模拟，获得平均电流。可以将模拟得到的平均电流缩减 20% ~25%，具体大小应考虑工艺的电迁移规则的严格性。最后，采用确定的电流值定义每层金属主电源线的最小宽度。这是一个稳妥的方式，因为它假定所有的电流从顶层的金属流向下面的电路并且没有横向电流产生。

4.12 集成问题

在较大的数字 IC 内部集成混合信号模块会变得困难，并且最终必须作出某些程度上的折衷。本节将讨论混合信号模块中的各种集成问题，并就如何确定各种选项的潜在影响方面给出指导意见。

4.12.1 芯片四角区域的影响

在给定了必须考虑的所有因素后，在较大的数字模块中集成模拟和混合信号电路会变得困难。以一个在其关键的定时部分采用了一个 DLL 单元的存储器电路为例，大多数情况下都希望能使用相同的布局规划可以生成具有不同存储容量

的电路。但是，这就使得把 DLL 嵌入到核心电路中变得很困难。合理的布局位置是在芯片的四角，但是四角是应力最高的区域。应力会引起 10mV 或者更大阈值电压漂移。由此引出的一个较大潜在问题是，如果芯片尺寸较小，阈值电压偏移几乎观察不到；但是对同一产品系列中芯片尺寸较大的产品，与芯片尺寸大小密切相关的阈值电压漂移将会明显，导致性能的变化与芯片尺寸直接相关，而这是不希望的。图 4-32 显示的是一个大芯片的四角区域简化示意图。最敏感的电路应尽可能的放在远离芯片四角的地方，而将最不敏感的电路放在最靠近芯片四角的位置。对 PLL，最不敏感的可能是环路滤波器（假定阈值电压漂移不会对电容产生严重影响）。

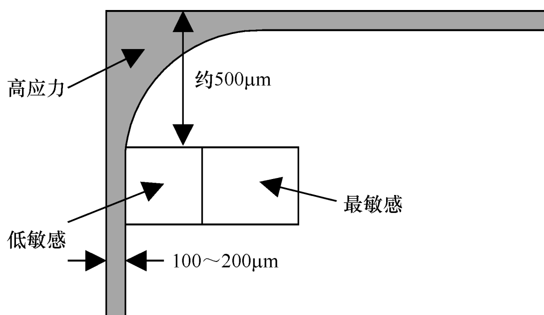


图 4-32　一个较大芯片（1.8cm×1.8cm）的右上角，图中标示了最大应力区域

4.12.2　邻近电路的影响

把模拟和混合信号模块布局规划进一个器件时，必须考虑其邻近的电路。如果邻近电路的开关电流极大，将给衬底注入噪声、干扰局部电源电压并引入较大的温度梯度，从而会严重影响模拟和混合信号电路的性能。衬底噪声已经讨论过了，下面只讨论其他两种效应。

模拟和混合信号电路中每个模块对温度和电源电压的灵敏度因子都可以被确定。例如 PLL 中高速计数器的这两个灵敏度因子分别为 0.25ps/℃ 和 0.4ps/mV。通过首先确定电路对这些因素的灵敏度，就可以得到一个明智的决定以确定如何完成单元的布局规划。如果一个特定的单元对温度梯度不敏感，将它放置在能产生较大温度梯度的单元旁边对该单元不会产生什么影响。对于局部电源电压对单元的影响问题可以按照同一思路处理。

4.13　总结

在深亚微米工艺下设计模拟和混合信号电路存在前面所讨论的各种问题。为了确保设计成功，设计者必须理解与工艺相关的问题。为了确保仿真能精确预测实际电路的性能，关键问题是知道并理解模型中包括哪些参数。即使设计正在生产线实现，为了保证在工艺被完全定义之前能确定潜在的问题，参与工艺开发正变得更加重要。

在项目中应该尽可能早的分析合适的结构，以研究尽可能多的选择。目前对片上系统集成电路中必须采用的许多混合信号电路，可以采用薄栅氧器件，至少在 90nm 工艺是这样。但是现在尚不清楚在 65nm 工艺节点是否还能继续使用。当有了一个可供仿真的基本电路图后，什么时候能提供可用的器件模型，是在完成最终设计之前实现结构折衷的关键问题。

在最初的架构阶段，必须考虑设计的各个方面问题，包括布局规划、主电源线、噪声隔离和去耦。在设计被执行之前，通过仔细策划设计的各个方面可以避免潜在的设计失误。

参 考 文 献

[1] R. A. Bianchi, G Bouche, and O. Roux-dit-Buisson, Accurate modeling of trench isolation induced mechanical stress effects on MOSFET electrical performance, *IEDM '02 Digest, Electron Device Meeting*, pp. 117–120, 2002.

[2] C. H. Choi, K. Y. Nam, Z. Yu, and R. W. Dutton, Impact of gate tunneling current in scaled MOS on circuit performance: a simulation study, *IEEE Trans. Electron Devices*, Vol. 48, No. 12, Dec. 2001.

[3] G. Chen, M. F. Li, C. H. Ang, J. Z. Zheng, and D. L. Kwong, Dynamic NBTI of p-MOS transistors and its impact on MOSFET scaling, *Electron Device Lett.*, Vol. 23, No. 12, pp. 734–736, Dec. 2002.

[4] T. C. Choi, R. T. Kaneshiro, R. W. Brodersen, P. R. Gray, W. B. Jett, and M. Wilcox, High-frequency CMOS switched-capacitor filters for communications application, *IEEE J. Solid-State Circuits*, Vol. 18, pp. 652–664, Dec. 1983.

[5] S. Yan and E. S. Sinencio, Low voltage analog circuit design techniques: a tutorial, *IEICE Trans. Analog Integrat. Circuits Syst.*, Vol. E00-A, No. 2, pp. 1–17, Feb. 2000.

[6] T. Serrano and B. Linares-Barranco, The active-input regulated-cascode current mirror, *IEEE Trans. Circuits Syst. II: Analog Digital Signal Process.*, Vol. 41, pp. 464–467, June 1994.

[7] K. N. Leung and P. K. T. Mok, A sub-1-V 15 ppm/°C CMOS bandgap voltage reference without requiring low threshold voltage device, *IEEE J. Solid-State Circuits*, Vol. 37, No. 4, pp. 526–530, Apr. 2002.

[8] T. Stockstad and H. Yoshizawa, A 0.9-V 0.5-μA rail-to-rail CMOS operational amplifier, *IEEE J. Solid-State Circuits*, Vol. 37, No. 3, pp. 286–292, Mar. 2002.

[9] J. F. Duque-Carrillo, J. L. Ausin, G. Torelli, J. M. Valverde, and M. A. Dominguez, 1-V rail-to-rail operational amplifiers in standard CMOS technology, *IEEE J. Solid-State Circuits*, Vol. 35, No. 1, pp. 33–44, Jan. 2000.

[10] J. M. Carrillo, J. F. Duque-Carrillo, G. Torelli, and J. L. Ausin, Constant-g_m constant-slew-rate high-bandwidth low-voltage rail-to-rail CMOS input stage for VLSI cell libraries, *IEEE J. Solid-State Circuits*, Vol. 38, No. 8, pp. 1364–1372, Aug. 2003.

[11] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, Cambridge University Press, Cambridge, 2000.

[12] J. G. Maneatis, Low-jitter process-independent DLL and PLL based on self-biased techniques, *IEEE J. Solid-State Circuits*, Vol. 31, No. 11, pp. 1723–1732, Nov. 1996.

[13] J. Maget, M. Tiebout, and R. Kraus, Influence of novel MOS varactor on the performance of a fully integrated UMTS VCO in standard 0.25 μm CMOS technology, *IEEE J. Solid-State Circuits*, Vol. 37, No. 7, pp. 1–6, July 2002.

[14] D. Coolbaugh, E. Eshun, R. Groves, D. Haranel, J. Johnson, A. Harnmad, Z. He, V. Ramachandran, K. Stein, S. St. Ongel, S. Subbanna, D. Wang, R. Volant, X. Wang, and K. Watson, Advanced passive devices for enhanced integrated RF circuit performance, *Technical Digest, 2002 IEEE MTT-S International Microwave Symposium*, pp. 187–191, Seattle, WA, June 2002.

[15] R. Aparicio and A. Hajimiri, Capacity limits and matching properties of integrated capacitors, *IEEE J. Solid-State Circuits*, Vol. 37, No. 3, pp. 384–393, Mar. 2002.

[16] H. Samavati, A. Hajimiri, A. R. Shahani, G. N. Nasserbakht, and T. H. Lee, Fractal capacitors, *IEEE J. Solid-State Circuits*, Vol. 33, No. 12, pp. 2035–2041, Dec. 1998.

[17] W.-T. Wang, M.-D. Ker, M.-C. Chiang, and C.-H. Chen, Level shifters for high-speed 1-V to 3.3-V interfaces in a 0.13 μm Cu-interconnection/low-κ CMOS technology, *Proceedings of Technical Papers, VLSI Technology, Systems, and Applications International Symposium*, pp. 307–310, Apr. 18–20, 2001.

[18] M. Pfost and H.-M. Rein, Modeling and measurement of substrate coupling in Si-bipolar ICs up to 40 GHz, *IEEE J. Solid-State Circuits*, Vol. 33, pp. 582–591, Apr. 1998.

[19] H. M. Chen, M. H. Wu, L. Chang, and C. F. Wu, The study of substrate noise and noise-rejection-efficiency of guard-ring monolithic integrated circuits, pp. 123–128, *Electromagnetic Compatibility*, 2000, IEEE International Symposium, Volume 1, 21–25 Aug 2000.

[20] J. P. Z. Lee, F. Wang, A. Phanse, and L. C. Smith, Substrate cross talk noise characterization and prevention in 0.35 μm CMOS technology, *IEEE Custom Integrated Circuits Conference*, pp. 479–482, 1999.

[21] A. Samavedam, A. Sadate, K. Mayaram, and T. S. Fiez, A scalable substrate noise coupling model for design of mixed-signal ICs, *IEEE J. Solid-State Circuits*, Vol. 35, pp. 895–904, June 2000.

[22] X. Aragones and A. Rubio, Experimental comparison of substrate noise coupling using different wafer types, *IEEE J. Solid-State Circuits*, Vol. 34, pp. 1405–1409, 1999.

[23] D. K. Su, M. L. Loinaz, S. Masui, and B. A. Wooley, Experimental results and modeling techniques for substrate noise in mixed-signal integrated circuits, *IEEE J. Solid-State Circuits*, Vol. 28, pp. 420–430, Apr. 1993.

[24] S. Masui, Simulation of substrate coupling in mixed-signal MOS circuits, *Digest of Technical Papers, IEEE Symposium on VLSI Circuits*, pp. 42–43, 1992.

[25] M. J. Chen, C. Y. Huang, P. N. Tseng, N. S. Tsai, and C. Y. Wu, Design model for minority-carrier well-type guard-rings in CMOS circuits, *IEEE Custom Integrated Circuits Conference*, pp. 4.5.1–4.5.4, 1991.

[26] R. Singh and S. Sali, Substrate noise issues in mixed-signal chip designs using SPICE, *10th International Conference on Electromagnetic Compatibility*, pp. 108–112, Sept. 1–3 1997.

[27] M. Ingels and M. S. J. Steyaert, Design strategies and decoupling techniques for

reducing the effects of electrical interference in mixed-mode ICs, *IEEE J. Solid-State Circuits*, Vol. 32, pp. 1136–1141, July 1997.

[28] K. M. Fukuda, T. Kikuchi, and M. Hotta, Measurement of digital noise in mixed-signal integrated circuits, *IEEE J. Solid-State Circuits*, Vol. 30, pp. 87–92, Feb. 1995.

[29] R. Gharpurey and R. G. Meyer, Modeling and analysis of substrate coupling in integrated circuits, *IEEE J. Solid-State Circuits*, Vol. 31, pp. 344–353, Mar. 1996.

[30] A. J. Rainal, Eliminating inductive noise of external chip interconnections, *IEEE J. Solid-State Circuits*, Vol. 29, pp. 126–129, Feb. 1994.

[31] K. Joardar, A simple approach to modeling cross-talk in integrated circuits, *IEEE J. Solid-State Circuits*, Vol. 29, pp. 1212–1219, Oct. 1994.

[32] B. R. Stanisic, R. A. Rutenbar, and L. R. Carley, Addressing noise decoupling in mixed-signal ICs: power distribution design and cell customization, *IEEE J. Solid-State Circuits*, Vol. 30, pp. 321–326, Mar. 1995.

[33] B. R. Stanisic, N. K. Verghese, R. A. Rutenbar, L. R. Carley, and D. J. Allstot, Addressing substrate coupling in mixed-mode ICs: simulation and power distribution synthesis, *IEEE J. Solid-State Circuits*, Vol. 29, pp. 226–238, Mar. 1994.

[34] N. K. Verghese, D. J. Allstot, and M. A. Wolfe, Verification techniques for substrate coupling and their application to mixed-signal IC design, *IEEE J. Solid-State Circuits*, Vol. 31, pp. 354–365, Mar. 1996.

[35] M. Pfost, H. M. Rein, and T. Holzwarth, Modeling substrate effects in the design of high-speed Si-bipolar ICs, *IEEE J. Solid-State Circuits*, Vol. 31, pp. 1493–1501, Oct. 1996.

[36] D. M. Monticelli, A quad CMOS single-supply op amp with rail-to-rail output swing, *IEEE J. Solid-State Circuits*, Vol. 21, pp. 1026–1034, Dec. 1986.

[37] P. Smeys, P. B. Griffin, Z. U. Rek, I. D. Wolf, and K. C. Saraswat, Influence of process-induced stress on devices characteristics and its impact on scaled device performance, *IEEE Trans. Electron Devices*, Vol. 46, pp. 1245–1252, June 1999.

[38] P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, Wiley, New York, 1993.

第 5 章　静电放电保护设计

5.1　引言

在集成电路（IC）中引入静电放电（Electrostatic Discharge，ESD）保护已有 30 多年了。伴随着工艺尺寸的缩小，ESD 设计面临着三个问题。第一，核心器件的栅氧化层不断变薄，使氧化层击穿电压减小到了一个难以保护的水平。如图 5-1 所示，薄栅氧器件的击穿电压已经降到小于 5V，而工作电压高达 1.5V，这就要求 ESD 保护器件必须工作在这两个电压之间。由于这两个界限如此接近，设计将变得更加困难。第二个效应是 I/O 速度前所未有的增快，由大容性负载要求的 I/O 缓冲 ESD 保护器件将会很大，不切实际。最后一个效应是多电源电压的引入，使得整个 ESD 保护方案更加复杂。一些电源可能只有极轻的负载，使得 ESD 保护很难足够快的开启以保护一些器件不受损伤。本章将介绍在几代工艺中使用的并且在 90nm 工艺以及更小尺寸工艺中仍将得到使用的几种 ESD 保护方案。

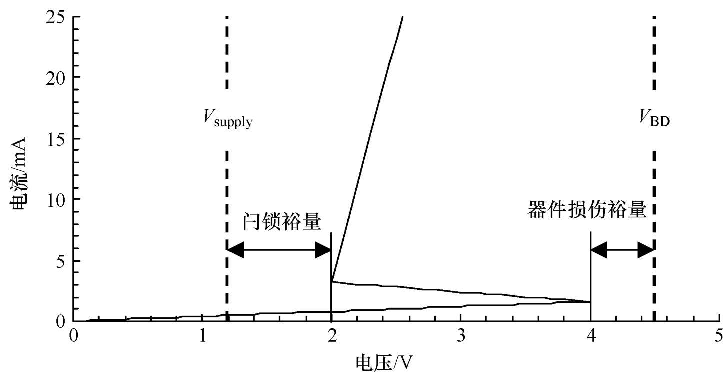


图 5-1　深亚微米工艺中 ESD 保护所面临的缩放问题

5.2　ESD 标准与模型

ESD 现象经常发生在两个或多个具有不同静电势的目标之间。对采用先进的深亚微米半导体工艺技术制造的 IC 产品，ESD 现象是影响 IC 安全工作的一个严重问题。在尺寸缩小的 CMOS 工艺中，采用更浅的结深、更薄的栅氧、轻掺杂

漏（LDD）结构和硅化物扩散的 MOS 器件，工作速度更高，功耗更低，但是它们抵御 ESD 应力的能力也更弱。器件被 ESD 损伤是由于受到快速产生的热量或快速产生的强电场作用。为了预测 IC 抗 ESD 的水平或查找对 ESD 敏感的位置，有几个组织制定了 ESD 标准。它们是 ESDA（静电放电协会）、AEC（汽车电子设备委员会）、EIA/JEDEC（电子工业联盟/电子器件工程联合协会）和 MIL－STD（美国军用标准）。对于元器件级别的 IC 产品，主要的测试方式是 HBM（人体模型）[1]，MM（机器模型）[2] 和 CDM（元件充电模型）[3]。

5.3　ESD 保护设计

5.3.1　ESD 保护方案

图 5-2 所示的片上保护设计概念被用来避免来自几乎随机组合的引脚之间的 HBM/MM ESD 应力损伤。对每一个输入或输出引脚，在压焊点与 VDD 和 VSS 电源线之间采用 ESD 钳位器件，用于泄放 I/O 引脚上的四种模式的 ESD 应力。为了克服引脚到引脚之间的 ESD 应力和 VDD 到 VSS 之间的 ESD 应力，在 IC 的 VDD 与 VSS 电源线之间必须放置电源轨 ESD 钳位电路[4,5]。对大多数逻辑 IC，I/O 电路的电源引脚通常与核心电路的电源引脚分开，以避免噪声耦合，并且可以减小地弹。图 5-3 所示为典型的采用独立电源引脚的整个芯片 ESD 保护方案。除了输入与输出端口的 ESD 钳位器件，对 IC 中所有器件和电路防止 ESD 损伤（特别是针对引脚到引脚和 VDD 到 VSS 的 ESD 应力），实现整个芯片保护的最重要的设计是合理排布电源线，并且在不同的电源线之间采用电源轨 ESD 钳位电路。如图 5-3 中虚线所示，ESD 电流泄放通路或在引脚对引脚 ESD 放电下的 IC，可以通过使用 I/O ESD 器件、金属电源线和电源轨 ESD 钳位电路来建立。只有

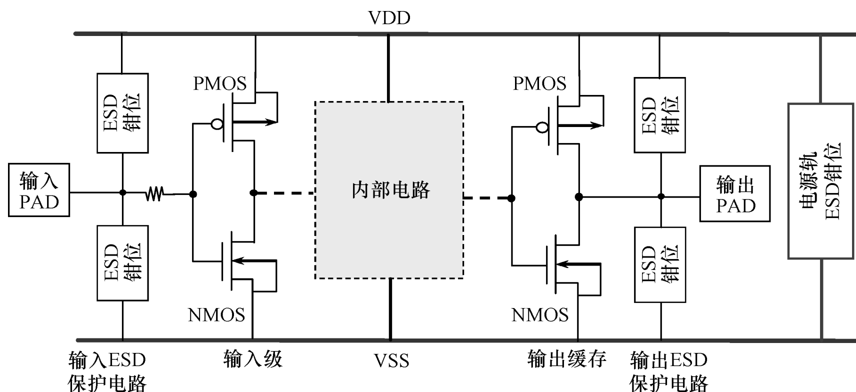


图 5-2　片上 ESD 保护设计的概念

使用正确的整片 ESD 保护计划，通过位于 I/O 端口的 ESD 钳位器件和位于电源轨之间的 ESD 钳位电路，才能实现对核心电路的有效保护。

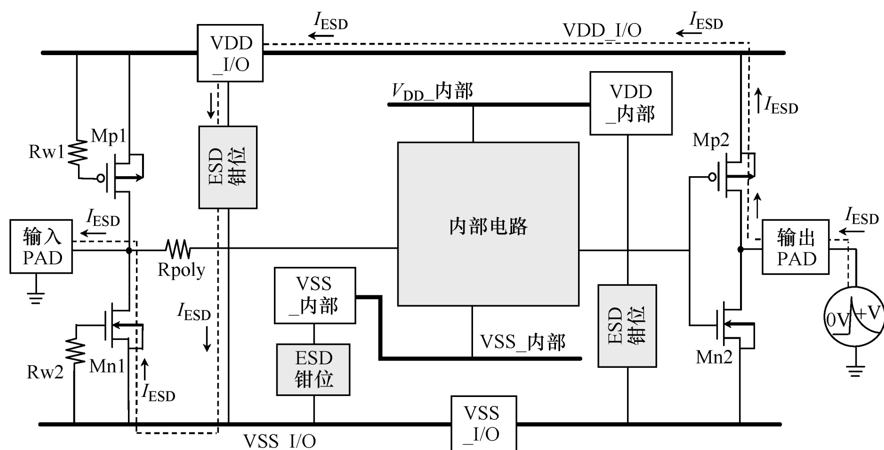


图 5-3　IC 中 I/O 电路和内部电路采用不同电源引脚的整片 ESD 保护方案

大多数传统的 ESD 保护设计集中在针对 I/O 端口的 ESD 保护电路和器件，这可以给四种引脚组合模式的 I/O 端口提供针对 ESD 应力的保护。但是，位于 I/O 端口的 ESD 保护电路或器件不能针对引脚到引脚和 VDD 到 VSS 的 ESD 应力实现对内部电路足够的保护，从而经常发生 ESD 应力引起内部电路失效，但在 I/O 端口的 ESD 器件并没有发生失效。对于采用多个独立电源的 IC 来说，建议采用的整片 ESD 保护方案如图 5-4 所示，一个著名半导体生产线的设计规则中

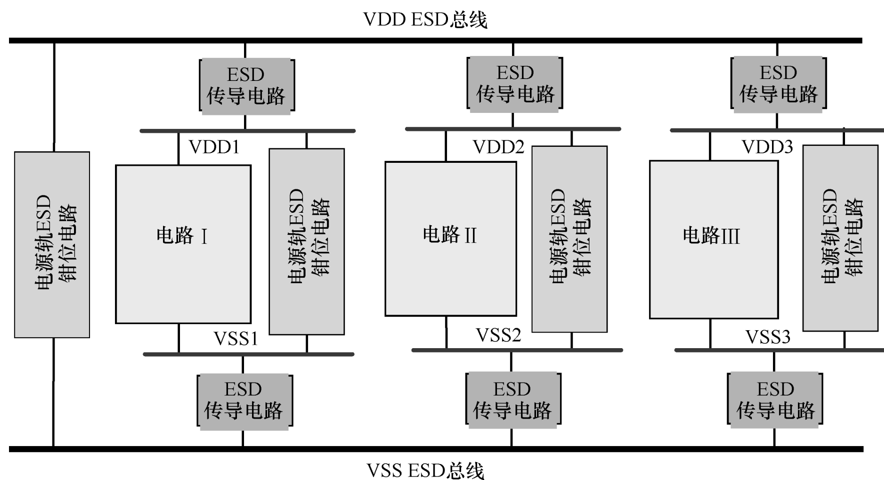


图 5-4　针对拥有多个不同电源线的 IC 整片 ESD 保护方案

采用了这种设计。VDD ESD 总线与独立电源（VDD1，VDD2，VDD3）之间的 ESD 传导电路使用层叠二极管[7]甚至是双向晶闸管（SCR）整流器[8,9]。对一个采用具有多个不同电平的复杂电源的 IC（例如片上系统（SoC）），采用多个 ESD 总线的整片 ESD 保护方案如图 5-5 所示。在整片 ESD 保护方案中，位于 I/O 端口和电源轨之间的 ESD 钳位器件仍然可以实现对内部电路的安全保护。

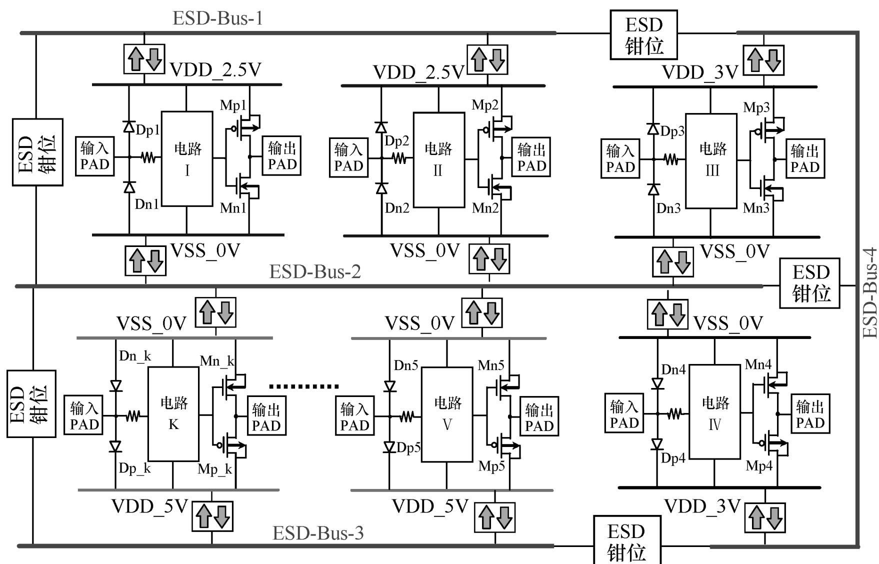


图 5-5　对有非常复杂的电源配置和多个不同电平的 IC，采用多 ESD 总线的全片 ESD 保护方案

5.3.2　ESD 保护器件开启的一致性

采用常规 CMOS 工艺可以实现的元器件有电阻、二极管、NMOS/PMOS、场氧器件（FOD）、纵向/横向双极型晶体管（BJT）、SCR 器件（p-n-p-n 结构）、电容和电感。因此可以采用这些器件组成 ESD 钳位器件或电路以实现要求的 ESD 保护水平（例如，2000V 的 HBM，200V 的 MM，1000V 的 CDM）。为了保持足够高的 ESD 级别，ESD 钳位电路中使用的器件应该有相应足够大的器件面积或尺寸以保证器件不被 ESD 引起的热量烧坏或损伤。然而，要在合理的硅片面积范围内绘制能维持足够高 ESD 级别的 ESD 钳位器件，必须采用版图优化设计技术[10-12]。对使用较小的版图面积实现足够高的 ESD 级别的 ESD 钳位器件的问题，参考资料［13-16］中给出了一些有效利用面积的先进设计方法。

在输出驱动中的 PMOS 和 NMOS 器件也经常作为 ESD 钳位器件以保护 I/O 引脚而不需要在端口增加额外的 ESD 钳位器件。为了给外部负载提供足够高的驱动电流并且能经受足够高的 ESD 应力，输出缓冲器中的 NMOS/PMOS 的沟道宽度一般设计为宽达几百微米。这种有很大器件面积的 ESD 保护器件一般采用多指梳状结构，以减小总的版图面积[17]。但是在 ESD 应力期间，起 ESD 保护作用的 MOSFET 多个梳齿并不能实现同时开启。如果只有其中部分 MOSFET“梳齿”开启，器件将会受到 ESD 放电的损伤[18]。这样，即使 MOSFET 有很大的器件面积，但是 ESD 保护电路的 ESD 级别并不高。为了提高多个“梳齿”开启的一致性，有报导采用栅驱动设计[19-22]和衬底触发设计[24-28]，以增加面积较大的 NMOS 器件的 ESD 鲁棒性。最近研究发现当栅压稍微增加时，栅驱动 NMOS 的鲁棒性将急剧下降[16,22]。这是因为栅驱动设计引起了穿过 NMOS 强反型沟道的较大放电电流[29]；因此 NMOS 器件很轻易被 ESD 能量烧毁。栅驱动和衬底驱动技术能改善大面积 ESD 保护器件的鲁棒性。但是从能带图的角度看，较高的栅偏置将会导致较大的沟道电流和较高的栅氧电场，并损伤 MOSFET[29]。这一效应引起了栅驱动器件 ESD 鲁棒性的衰退。与栅驱动设计相比，衬底驱动设计能避免沟道电流的形成并且能增强空间电荷区使得较高的 ESD 电流远离沟道表面。因此，在纳米级 CMOS 技术中，衬底驱动设计是改善 CMOS 器件 ESD 鲁棒性的最有效方法之一[30]。

5.3.3　ESD 注入与金属硅化物阻挡

除了用于改善有限版图面积内 ESD 钳位器件的 ESD 鲁棒性的版图设计或触发技术外，还开发出了一些需要采用附加掩膜层的工艺调整技术来进一步改善 I/O 端口的 ESD 鲁棒性。为了提高这些钳位器件的 ESD 鲁棒性，有报导在工艺流程中引入 ESD 注入以改变 ESD 保护器件的结构[31-37]。采用 N 型 ESD 注入覆盖 LDD 的峰值结构部分，并且在起 ESD 保护作用的 NMOS 器件中形成更深的结[31,32]。采用 NMOS 漏结下的 P 型 ESD 注入，可以减小反向结击穿电压并且更早的开启 NMOS 器件中的寄生横向双极器件[33,34]。掺杂浓度较高的 P 型 ESD 注入还可以用来减小二极管或场氧器件的反向结击穿电压，使得器件在反向偏置条件下经受更高的 ESD 应力[35]。此外，N 型和 P 型 ESD 注入都可以在 NMOS 器件中使用，以保证更高的 ESD 鲁棒性[36]。目前已有在同一 CMOS 工艺中采用各种不同 ESD 保护方案之间的实验结果对比[37]。为了改善 NMOS，特别是 MM ESD 鲁棒性，已有关于改进 ESD 注入设计的报道[38,39]。

改善 I/O 器件 ESD 鲁棒性的另一种工艺调整是使用附加掩膜层来阻挡连接到 I/O 压焊点的器件的漏区形成金属硅化物的扩散[40-46]。一个相似的但

不需要附加掩膜层的方法是在 NMOS 的漏区形成一个 n 阱，阻挡 I/O 器件的漏区形成金属硅化物扩散，以改善 ESD 鲁棒性[47,48]。最近，也有报导采用哑元栅来阻挡 I/O 器件漏区形成金属硅化物的扩散而不需要使用额外的掩膜层[49,50]。

5.3.4　ESD 保护指导意见

下面是设计 I/O 引脚或者电源引脚的 ESD 钳位器件或电路应遵循的一些指导意见：

（1）给 IC 提供有效的 ESD 保护方案，当 IC 受到 ESD 应力作用时，能旁路 ESD 应力。

（2）在正常工作条件下，IC 能传送正常的 I/O 信号并保持 ESD 处于非激活状态。

（3）减小输入电容和电阻（尽可能使其最小），使 I/O 信号延迟在可接受的范围。

（4）在合理的版图面积（尽可能小）内具有高 ESD 鲁棒性。

（5）在 CMOS IC 中维持一个较高的抗闩锁能力（所有的 I/O 器件和 ESD 保护器件必须采用保护环包围）。这种 ESD 保护器件的制造应该与工艺技术兼容（如果可能的话，不需要附加的掩膜层，不要求调整工艺步骤）。

（6）工作时，ESD 电路不能影响 I/O 电路的功能（例如应用高电压容限的 I/O 或断电工作）。

5.4　针对高速 I/O 的低电容 ESD 保护设计

5.4.1　高速 I/O 或模拟引脚的 ESD 保护

数字输入引脚通常采用的两级结构 ESD 保护设计，如图 5-6 所示。其中使用了一个栅极接地的短沟 NMOS 作为第二级保护器件，它可以钳位输入电路栅氧上的过应力电压。为了提供高 ESD 保护水平，第一级保护级采用一个鲁棒器件（例如 SCR、场氧器件或长沟道 NMOS）作为主要泄放单元以旁路输入压焊点上的 ESD 电流。在输入 ESD 保护电路的第一级与第二级之间，增加一个电阻来限制流过第二级短沟 NMOS 的 ESD 电流。其阻值取决于第一级 ESD 钳位器件的开启电压和第二级短沟 NMOS 器件的二次击穿电流。在第二级栅接地 NMOS（GGNMOS）被过应力 ESD 电流损伤前，第一级 ESD 钳位器件必须被触发开启以旁路 ESD 电流。如果第一级 ESD 钳位器件的开启电压较高，则电阻应该足够大：大约在几百欧姆的级别。这种两级 ESD 保护设计能给数字输入引脚提供很高的

ESD 容限级别。但是较大的串联电阻和 ESD 钳位器件较大的结电容会引起较长的输入信号 RC 延迟，因此不适合用于模拟信号和高速 I/O 应用，而这两种情况在更先进的工艺中变得更占主导地位。

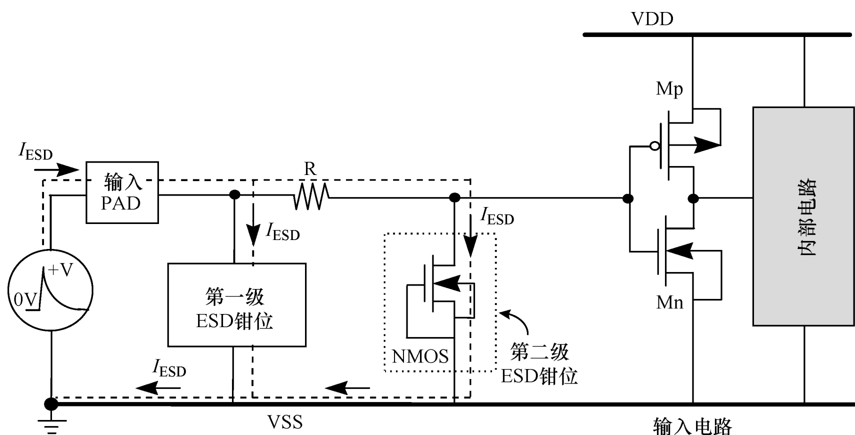


图 5-6　CMOS IC 中数字输入引脚通常采用的两级结构 ESD 保护电路

对模拟输入信号和高速 I/O 应用，在输入压焊点与内部电路之间一般不允许存在串联电阻。因此图 5-6 所示两级 ESD 保护设计不再适合于模拟和高速 I/O 引脚。为了保护模拟或高速 I/O 引脚，一般采用如图 5-7a 所示的只有单级 ESD 设计的 ESD 保护电路，其中采用 GGNMOS 作为 ESD 钳位器件。在先进的深亚微米 CMOS 技术中，由于缺少了限制 ESD 电流流向 GGNMOS 的串联电阻，也缺少 NMOS 器件的 ESD 鲁棒性，因此使 ESD 保护严重衰退[51,52]。GGNMOS 的设计一般具有很大的面积、漏与多晶硅栅接触之间的版图间距较宽，以维持能满足要求的 ESD 级别[16,52,53]。在深亚微米 CMOS 工艺中包括了附加的金属硅化物阻挡掩膜[40-44]以增加 ESD 钳位器件的鲁棒性。图 5-7b 是采用金属硅化物阻挡漏区 GGNMOS 的剖面示意图。但是有更大器件面积和更宽漏

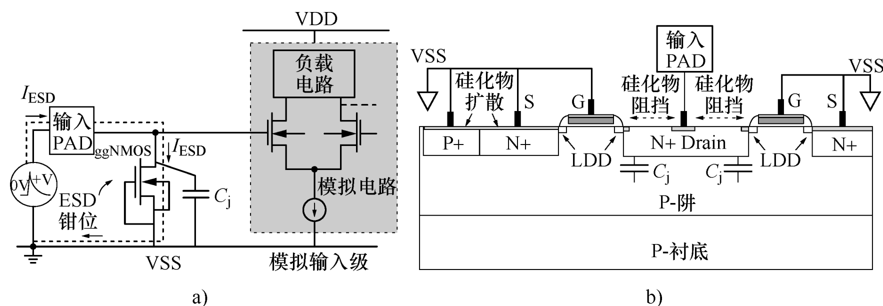


图 5-7　a）用于 CMOS IC 中模拟输入引脚的单级 ESD 饱和电路

b）具有金属硅化物阻挡漏区的 GGNMOS 的剖面图

区的 GGNMOS 给输入压焊点增加了很大的寄生漏电容。这样的寄生漏电容是非线性的并且电容大小与输入电压水平有关，对实际的高速 I/O 和模拟输入，这是不可接受的。

对一些高分辨率的数据转换器电路，要求在输入电压范围内模拟信号输入脚的输入电容尽可能保持为常数。高速模拟电路中，特别是在单端输入应用中，主要的失真源是模拟信号输入压焊点的 ESD 钳位器件具有的与电压有关的非线性电容。参考文献［65］报道了由于输入 ESD 钳位器件的非线性电容造成的电路性能的典型衰退情况。对于输入电压从 0 到 2V 的变化，输入电容的变化范围为 4～2pF。电容的变化引起模数转换器（ADC）谐波失真的增加，并且使 ADC 的精度从 14 位退化到 10 位。这样，在采用缩小尺寸 CMOS 技术的高精度模拟应用中，设计有效的 ESD 保护电路已面临挑战。5.4.2 节将详细介绍具有较小输入电容、无串联电阻和高 ESD 级别优势的小电容 ESD 保护设计。

5.4.2 小电容 ESD 保护设计

高速 I/O 或模拟信号引脚的小电容 ESD 保护电路如图 5-8a 所示。采用金属硅化物的 0.35μm 工艺 CMOS 单元库的这种小电容 ESD 保护电路参考版图如图 5-8b 所示。在图 5-8a 中，Dp1（Dn1）是 Mp1（Mn1）漏区的寄生结二极管。为了减小高速 I/O 或模拟信号引脚的输入电容，Mn1 和 Mp1 都设计为面积很小的器件：宽长比（W/L）仅 50/0.5（μm/μm）。采用典型 0.35μm 金属硅化物 CMOS 工艺中的宽长比（W/L）为 50/0.5（μm/μm）的 NMOS 器件，在 PS 模式 ESD 应力下放电时，其 HBM ESD 水平小于 500V。但是同样大小和工艺的 NMOS，在 NS 模式应力放电时，却可以经受 8000V 的 HBM ESD 应力。在 PS 模式（NS 模式）ESD 应力下，NMOS 工作在漏击穿条件下（漏二极管正向导通），能旁路 ESD 电流。ESD 钳位器件的功耗等于 ESD 放电期间 ESD 电流与器件工作电压的乘积。因此，NMOS 在 PS 和 NS 模式下的 ESD 应力水平差别极大。同样，器件面积较小的 PMOS 在 PD 模式下的 ESD 应力级别较高，而在 ND 模式应力下 ESD 应力级别则低得多。

为了避免小面积的 Mn1 和 Mp1 在 PS 和 ND 模式下进入漏击穿条件，在模拟电路 ESD 保护电路的电源轨之间同时设计了一个有效的 ESD 钳位电路，以增加总体 ESD 保护级别。在图 5-8a 中，当压焊点上有 PS 或 ND 模式的 ESD 应力放电时，基于 RC 的 ESD 检测电路[5]被用来触发开启器件 Mn3。图 5-9a 和图 5-9b 中的虚线部分分别表示 PS 和 ND 模式放电时模拟电路 ESD 保护电路模拟信号引脚的 ESD 电流路径。因为在 PS 模式（Mp1 在 ND 模式）应力下 Mn1 并没有工作在漏击穿条件下，ESD 电流通过 Mp1 中的正向偏置二极管 Dp1（Mn1 中的 Dn1）

a)

b)

图 5-8　a）用于高速 I/O 或模拟信号引脚的小电容 ESD 保护电路　b）采用 0.35μm 金属硅化物 CMOS 工艺的高速 I/O 或模拟信号引脚的小电容 ESD 保护电路版图实例

和开启的 Mn3 旁路流出。Mn3 的尺寸被特意设计得更大（在图 5-8b 中 W/L = 1880μm/0.5μm）以维持高 ESD 级别。虽然大尺寸的 Mn3 有较大的结电容，但是这个电容对模拟信号压焊点并没有影响。因此，模拟信号引脚能在四种模式的 ESD 应力下维持更高的 ESD 级别而输入电容却较小。在输入脚保护器件和电源 ESD 保护器件之间的寄生电阻很关键，必须做到最小化，以确保器件在 ESD 放电期间能快速开启，同时 IR 压降被最小化。

当输入引脚上出现模拟信号引脚对引脚 ESD 应力放电时，建议采用的这种保护电路中 ESD 电流路径如图 5-10 所示。在引脚对引脚 ESD 应力期间，IC 中的 VDDA 和 VSSA 电源线悬空。ESD 电流通过输入 ESD 保护电路 Mp1 中的结二极管 Dp1，从放电压焊点被引到 VDDA 电源线。因此，ESD 能量对 VDDA 线充电。VSSA 线的电平最初接近地电平，因为 VSSA 线通过另一个输入 ESD 保护电路中的 Mn4 的二极管 Dn4 连接到接地压焊点。两个不同的输入引脚上的 ESD 应

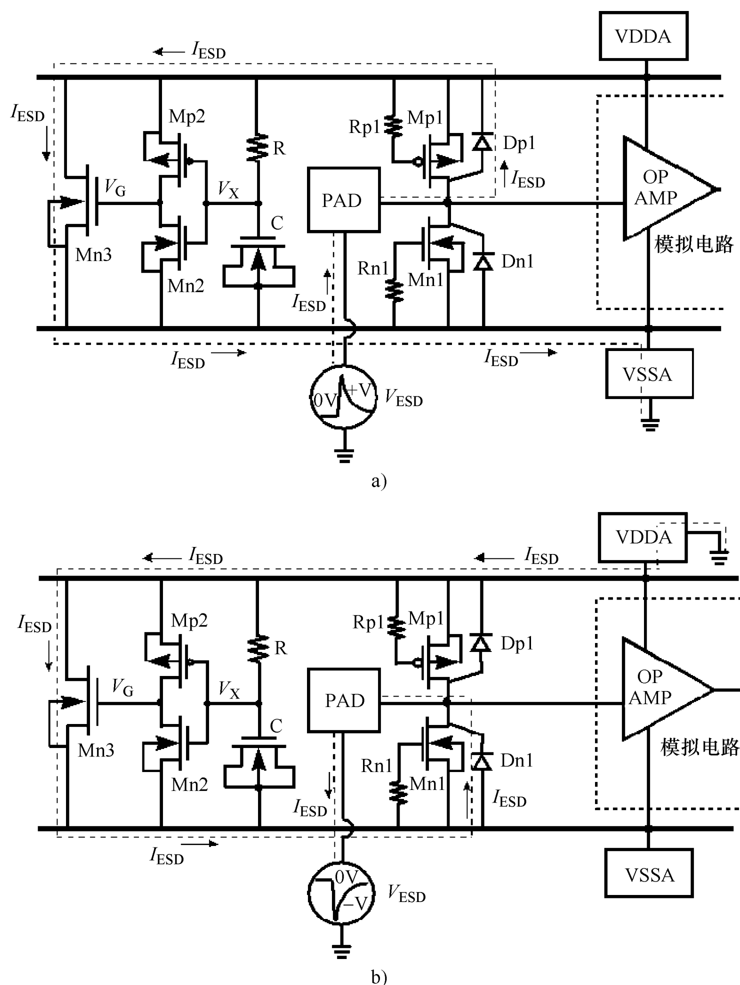


图 5-9　输入引脚在

a）PS 模式和　b）ND 模式 ESD 应力时低电容　ESD 保护电路的 ESD 电流路径

力电压加载到 VDDA 与 VSSA 电源线之间。连接在 VDDA 和 VSSA 电源线间的器件 Mn3 被基于 RC 的 ESD 检测电路开启，从而旁路了从 VDDA 到 VSSA 的 ESD 电流。最后，ESD 电流通过 Mn4 中正向偏置二极管 Dn4 从 VSSA 电源线流出芯片到接地压焊点。通过设计合适的 ESD 检测电路快速开启 Mn3[10]，引脚到引脚 ESD 应力可以通过差分输入级的栅氧被快速的泄放。通过这种设计，无须在输入压焊点和内部模拟电路之间增加任何的串联电阻就可以使模拟差分输入级的栅氧得到充分的保护。因此，由于小电容 ESD 保护电路的作用，从压焊点到内部电路，输入信号可以获得最宽的带宽。

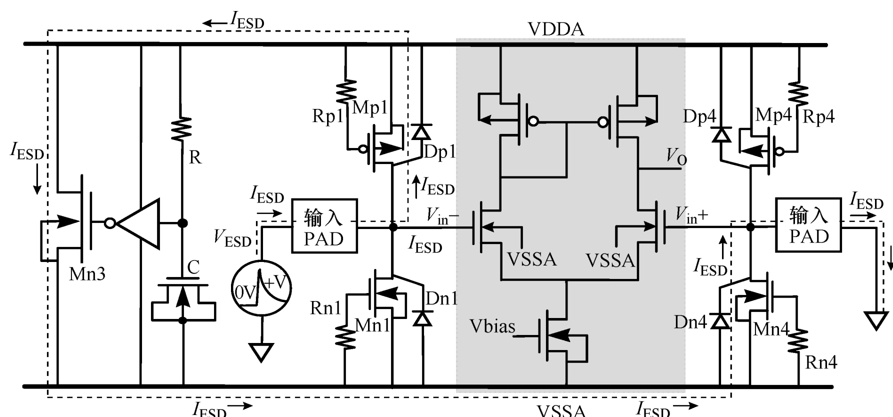


图 5-10　输入引脚存在模拟信号引脚对引脚 ESD 应力放电时，低电容 ESD 保护电路的 ESD 电流路径

5.4.3　输入电容的计算

用于高速 I/O 或模拟信号引脚的这种小电容 ESD 保护电路的输入电容可以由下式得出：

$$C_{in} = C_{pad} + C_n + C_p$$

式中，C_{pad}是键合区的寄生电容；C_p（C_n）是 Mp1（Mn1）的漏结电容和漏栅重叠电容。单个 NMOS 或 PMOS 的漏结电容与偏置有很强的依赖关系。当输入信号的电平不同时，采用图 5-7a 所示的单 NMOS 管的上述 ESD 保护设计的输入电容的变化范围将会很大。但是即使输入信号摆幅从 0 到 Vdd（3V）时，采用互补 NMOS 和 PMOS 的小电容 ESD 保护电路（图 5-8a）的输入电容也几乎保持为常数。器件面积不同的小电容 ESD 保护电路的总输入结电容可以通过 SRAR-HSPICE CAD 工具的频域函数 pin-capacitance-measurement 仿真[56]精确计算。

仿真结果如图 5-11 所示，其输入压焊点电压不同，Mn1 和 Mp1 的沟道长度为固定值 0.5μm，而沟道宽度从 50～400μm 变化。Mn1 与 Mp1 的漏接触孔到多晶硅栅的间距都为 3.4μm，而源极接触孔到多晶硅栅的间距为 1.55μm。对于 Mn1 和 Mp1 的器件面积都为 50/0.5(μm/μm) 的情况，输入电压从 0V 摆幅到 3V 时，小电容 ESD 保护电路的输入电容仅从 0.37pF 变化到 0.4pF。但是采用图 5-7a 所示 W/L＝400/0.5(μm/μm) 的 GGNMOS 的传统 ESD 保护电路，在输入电压从 0V 摆幅到 3V 时，其输入电容从 1.83pF 变为 1.12pF。

对 0.35μm CMOS 工艺的互连键合，金属压焊点的版图尺寸为(96×96)μm^2，相应的寄生电容值为 0.67pF。因此，即使输入信号有从 0V 到 3V 的电压摆幅，包括压焊点寄生电容在内的低电容 ESD 保护电路的总输入电容为 1.04pF 到

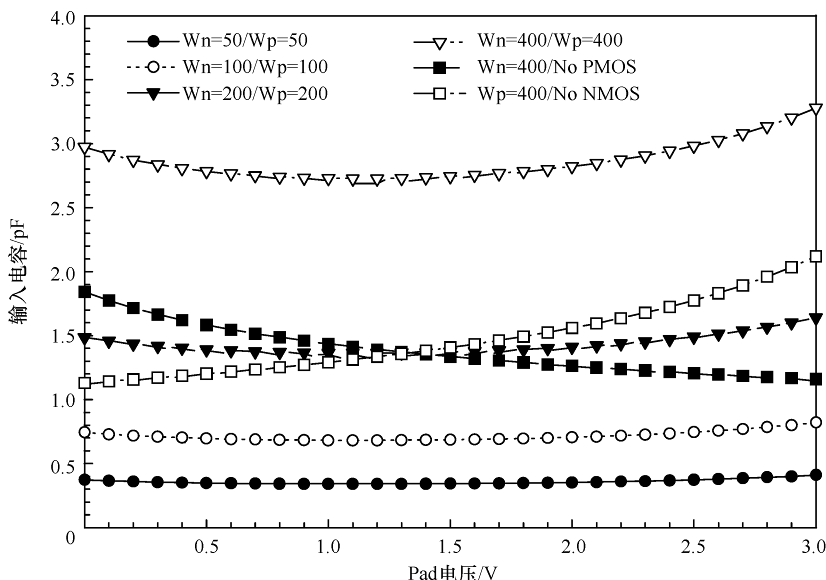


图 5-11　压焊点输入电平变化时，采用不同尺寸的 Mn1 和 Mp1 的低电容 ESD 保护电路的输入结电容

1.0pF。由于输入电容小并且几乎是常数，因此小电容 ESD 保护电路特别适合于模拟和数字信号引脚的高精度和高速 I/O 应用。为了进一步减小压焊点产生的寄生电容，已经开发出了小电容压焊点结构[57]。在 ESD 器件（Mn1 和 Mp1）与输入压焊点之间增加二极管可以使电容进一步减小。采用 0.25μm CMOS 工艺开发的层叠结构多晶硅二极管可以保护工作在 2.4GHz 的 RF 低噪声放大器（LNA）的 RF 输入脚[58]。

输入电容中的主要非线性源是图 5-5 所示的小电容 ESD 保护电路中的 Mn1 和 Mp1 漏区结电容，它们随偏置大小变化非常明显。当输入端口的输入信号电压增加时，Mn1 的漏结电容减小而 Mp1 的漏结电容增加。相反，当输入端口的输入信号电压减小时，Mn1 的漏结电容增加而 Mp1 的漏结电容减小。如果合理设计器件 Mn1 和 Mp1 的版图面积和间距，这种互补型低电容 ESD 保护电路的输入电容就可以几乎保持为常数。一个用于优化 ESD 钳位器件版图面积和间距的设计模型已经开发出来，可以保持低电容 ESD 保护电路的输入电容几乎为常数[70]。低电容 ESD 保护电路总输入电容的变化可以低于 1%。这个绝对误差考虑了 NMOS 和 PMOS 器件掺杂浓度的差异。杂质浓度的设定值和实际值之间的差异将导致非线性电容的增加。

5.4.4　ESD 鲁棒性

实际情况下，小电容 ESD 保护电路已经作为运算放大器的输入电路在

0.35μm 金属硅化物工艺中获得了运用。在测试芯片中，同相和反相输入引脚都采用了前面介绍的模拟电路 ESD 保护电路保护。在器件 Mn1，Mp1 和 Mn3 中采用了金属硅化物阻挡掩膜以改善它们的 ESD 鲁棒性，但是没有使用额外的 ESD 注入工艺调整。

对获得的模拟电路 ESD 保护电路，采用 ESD 模拟器产生的 HBM（人体模型）和 MM（机器模型）ESD 应力进行放电。表 5-1 中列出了最大 ESD 电压测试结果，包括模拟信号引脚对引脚 ESD 应力。失效标准定义为 ESD 放电后，在 5V 偏置电压下压焊点泄漏电流超过 1μA。如表 5-1 所示，低电容 ESD 保护电路能成功的给模拟信号引脚提供高于 6000V（400V）的 HBM（MM）ESD 保护而无须在压焊点与内部电路之间增加任何的串联电阻。

表 5-1　低电容 ESD 保护电路在 HBM 与 MM ESD 测试条件下的 ESD 级别

ESD 测试中的引脚组合					
	PS 模式	NS 模式	PD 模式	ND 模式	引脚到引脚
HBM/V	6000	−8000	7000	−7000	6000
MM/V	400	−400	400	−400	400

为了进行比较，在同一测试芯片中也设计了如图 5-7a 所示采用 W/L = 480/0.5（μm/μm）GGNMOS 的传统 ESD 保护电路，用于模拟信号输入脚。图 5-7a 中设计的 HBM PS 模式的 ESD 级别约为 3kV，但是它的模拟信号引脚对引脚 HBM ESD 级别低于 500V。引脚对引脚 ESD 损伤位置在运算放大器电路的第一输入级的多晶栅上。所以常规 ESD 保护设计在引脚对引脚 ESD 应力期间不能保护深亚微米 CMOS 技术中差分输入级较薄的栅氧化层。

虽然仅采用 0.35μm 工艺中制造了这个测试结构，但是它仍然适用于更先进的工艺。主要的差异是必须调整物理器件尺寸使得 NMOS 和 PMOS 器件的扩散电容相匹配。

5.4.5　开启验证

为了验证小电容 ESD 保护电路在引脚对引脚 ESD 应力下开启的有效性，在运算放大器的反相引脚施加脉冲发生器（HP 8118A）产生的一个方波电压脉冲，而运算放大器的同相管脚接地并且 VDDA 与 VSSA 悬空。在正和负引脚到引脚 ESD 应力条件下的验证实验电路图分别如图 5-12a 和图 5-12b 所示。

在正的引脚到引脚 ESD 应力条件下测量的电压波形如图 5-13a 和图 5-13b 所示。图 5-13a 所示的电压波形是直接由 HP 8118A 脉冲发生器产生的高度为

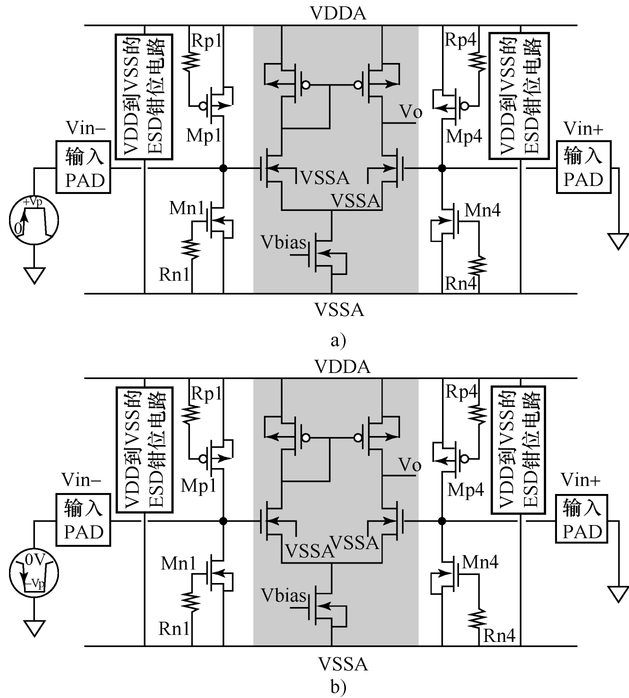


图 5-12 建立试验以验证低电容 ESD 保护电路在正（图 a）和负（图 b）的引脚到引脚 ESD 应力条件下开启的有效性

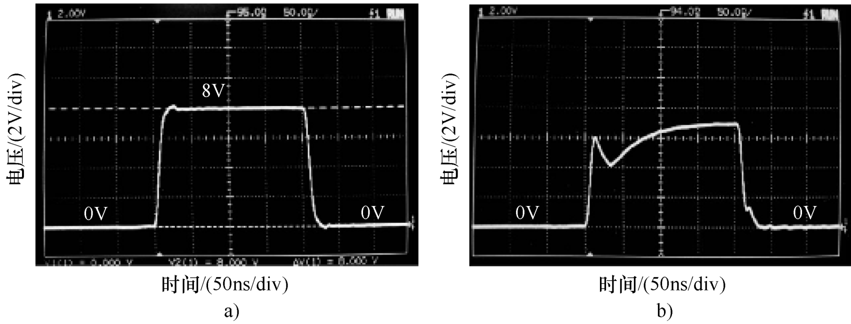


图 5-13 a）由脉冲发生器产生的 8V 脉冲电压测量波形 b）在图 5-12a 所示的引脚到引脚应力条件下，8V 电压脉冲被加到模拟反相输入引脚时的衰退波形

8V、脉宽为 200ns 的电压脉冲。8V 电压脉冲的上升时间约为 10ns，这与 HBM ESD 脉冲的上升时间相似。0.35μm 无附加 ESD 注入工艺调整的金属硅化物 CMOS 工艺中的 NMOS Mn1 的漏极击穿电压约为 8.5V。因此，脉冲高度为 8V 的电压脉冲不会引起模拟电路 ESD 保护电路中 Mn1 的漏极击穿。在模拟信号引脚加这样一个电压脉冲可以真实验证建议的模拟电路 ESD 保护电路开启的有效性。

虽然这个正电压脉冲被加到图 5-12a 所示的输入引脚，但是小电容 ESD 保护电路对他起到钳位作用，衰减的电压波形如图 5-13b 所示。在负的引脚对引脚 ESD 应力条件下测量得到的电压波形如图 5-14a 和图 5-14b 所示。图 5-14a 显示的电压波形是直接由脉冲发生器产生的脉冲高度为 －8V、脉宽为 200ns 的负电压脉冲。当该负电压脉冲加到如图 5-12b 所示的输入脚时，被模拟电路 ESD 保护电路钳位到仅 －5V 的水平，其衰退的电压波形如图 5-14b 所示。由图 5-13b 和 5-14b可见，在引脚到引脚 ESD 应力条件下，电压脉冲确实被低电容 ESD 保护电路钳位了。因此在引脚到引脚 ESD 应力条件下，位于输入级的较薄栅氧化层得到有效保护。

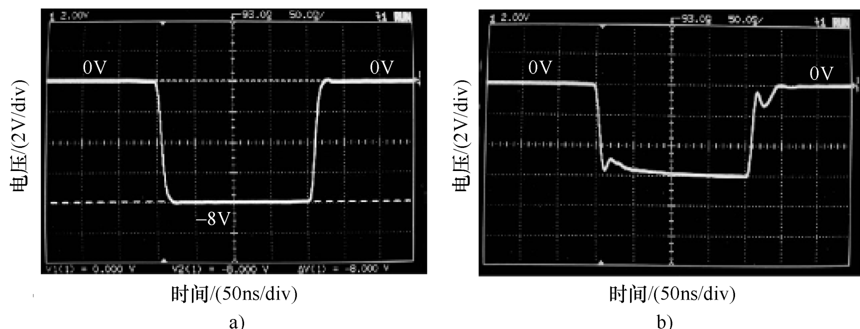


图 5-14　a）由脉冲发生器产生的 －8V 脉冲电压测量波形　b）在图 5-12b 所示的引脚到引脚应力条件下，－8V 电压脉冲被加到模拟反相输入引脚时的衰退波形

在 PS 和 ND 模式 ESD 应力下小电容 ESD 保护电路开启的有效性也得到了验证。PS 和 ND 模式的实验电路分别如图 5-15a 和图 5-15b 所示。当 PS 模式下电平为 8V 的一个电压脉冲（见 5-15a）加到输入脚时，小电容 ESD 保护电路将钳位这个正电压脉冲，压焊点上衰减了的波形如图 5-16a 所示。当 ND 模式下如图 5-15b 所示的负电压脉冲加到输入脚时，负电压脉冲（－8V）被小电容 ESD 保护电路钳位，压焊点上的衰退波形如图 5-16b 所示。这实际上验证了小电容 ESD 保护电路开启的有效性。

从上面的实验验证可知，通过 VDD 到 VSS 的 ESD 钳位器件 Mn3，正或负电压脉冲被小电容 ESD 保护电路所钳位。小电容 ESD 保护电路中的 Mn1 和 Mp1 工作在结二极管正向条件下而不是漏击穿条件下，因此，即使在 Mn1 和 Mp1 器件的面积比常规保护器件的面积更小时，提出的模拟电路 ESD 保护电路还能承受较高的 ESD 级别。压焊点和内部电路之间无串联电阻的小电容 ESD 保护可以安全地保护输入级的较薄栅氧化层。由于 Mn1 和 Mp1 的器件面积更小，在高频模拟 I/O 或高速数字 I/O 应用时，连接到压焊点上的总结电容可以更小。

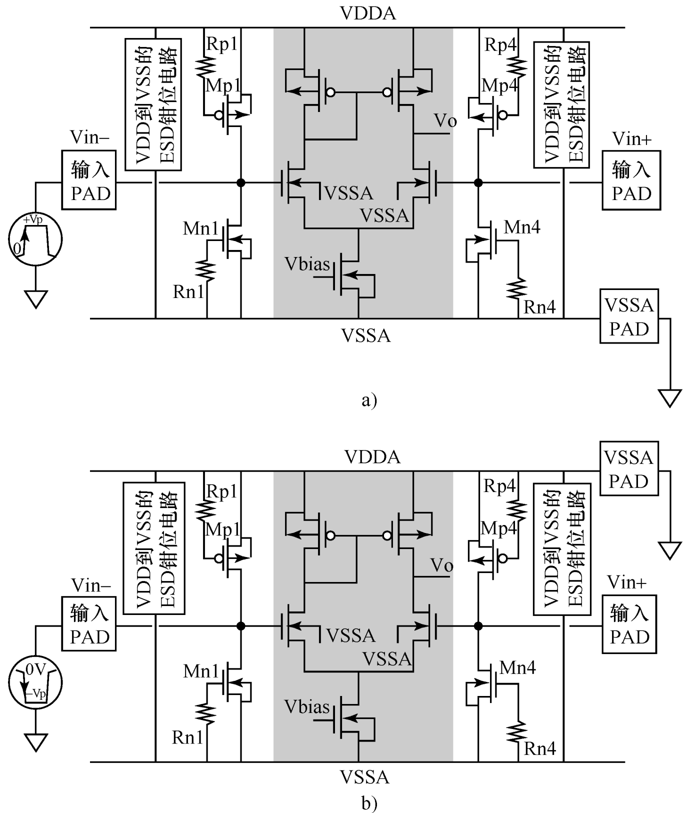


图 5-15 a）PS 模式 b）ND 模式 ESD 应力下验证低电容 ESD 保护电路开启有效性的实验电路

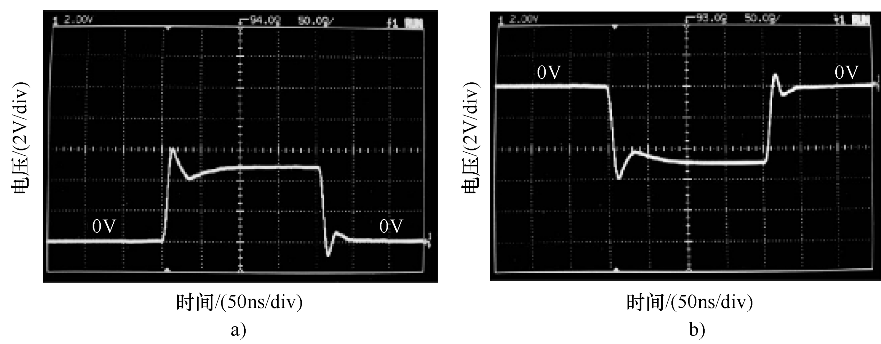


图 5-16 a）PS 模式 ESD 应力条件下，给输入引脚加正 8V 电压脉冲得到的衰退电压波形 b）ND 模式 ESD 应力条件下，给输入引脚加负 8V 电压脉冲得到的衰退电压波形

5.5 混合电压 I/O 的 ESD 保护设计

5.5.1 混合电压 I/O 接口

晶体管面积已缩减到纳米时代，导致电路的电源电压也快速减小。显然，更小的晶体管面积使芯片面积也更小，这就节省了硅材料的开销。更低的电源电压使功耗也更小（如果电路规模增大，更低的电源电压也使功耗的增加幅度减小）。在先进的现代 CMOS 工艺下，芯片设计将快速地转移到采用更低电压的设计，但是一些外围单元或其他 IC 还工作在较高电压（3.3V 或 5V）。换句话说，不同电源电压的芯片将并存于一个系统中。因为电源电压的这种混合，大多数微电子系统要求采用不同内部电源电压的半导体芯片或子系统的接口。由于不同电压电源的混合使用，必须正确设计芯片到芯片之间的接口 I/O 电路，以避免栅氧化层上出现过大的电应力[60]、避免输出器件中的热载流子衰退[61]、和阻止芯片之间出现不希望的泄漏电流[62,63]。例如，采用 CMOS 工艺实现的内部电源电压为 1.0 或 1.5V 的 IC，一般要求 3.3V 的接口。VDD 为 3.3V 的传统 CMOS I/O 缓冲器如图 5-17a 所示，其有一个输出级和一个输入级。当一个 5V 的信号加到 I/O压焊点时，输出 PMOS 和输出 PMOS 中寄生的漏阱结二极管提供了从 I/O 压焊点到 VDD 的泄漏电流通道（如图 5-17a 中虚线所示）。而且输出 NMOS 的栅氧化层、输入 ESD 保护的 GGNMOS 和输入反相级要经受 5V 输入信号的过应力作用。为了在不使用附加厚栅氧工艺（某些 CMOS 工艺中称为双栅氧工艺）的情况下解决栅氧可靠性问题，在混合电压 I/O 缓冲器中[66-72]，甚至在电源轨 ESD 钳位电路中[73]广泛使用层叠 MOS 配置。典型 3V/5V 容限混合电压 I/O 电路的基本电路如图 5-17b 所示[67]。当 5V 输入信号进入 I/O 压焊点后，连接 I/O 压焊点和 VDD 电源线的上拉 PMOS 的自偏置电路可以跟踪栅和 n 阱电压。实现栅跟踪功能和 n 阱自偏置电路模块的详细电路可参见参考文献［72］。

5.5.2 混合电压 I/O 接口的 ESD 问题

I/O 压焊点上的 ESD 应力有四种引脚组合模式：正脉冲作用于 VSS（PS 模式）、负脉冲作用于 VSS（NS 模式）、正脉冲作用于 VDD（PD 模式）和负脉冲作用于 VDD（ND 模式）四种 ESD 应力条件。为了使 CMOS 输出缓冲器有足够高的 ESD 鲁棒性，CMOS 缓冲器一般设计成具有大的器件面积和较宽的漏接触到多晶栅间距，这常常在 I/O 单元中占据较大的版图面积。在无需增加 I/O 单元器件面积的情况下，CMOS IC 电源线上的 VDD 到 VSS 的 ESD 钳位电路已成功用于改善 CMOS I/O 缓冲器的 ESD 鲁棒性[55]。在 PS 模式 ESD 应力条件下传统 CMOS

图 5-17　缓冲器典型电路框图

a）传统的 CMOS I/O 缓冲器　b）采用层叠 NMOS 和自偏置阱 PMOS 的混合电压 I/O 缓冲器

输出缓冲器的 ESD 电流通路如图 5-18a 中虚线所示，大部分 ESD 电流通过 PMOS 的寄生二极管和 VDD 到 VSS 的 ESD 钳位电路泄放到地。因此，传统的 CMOS 输出缓冲器可以经受高 ESD 应力。但是由于考虑到混合电压 I/O 缓冲器的泄漏电流问题，在 I/O 压焊点和 VDD 电源线之间将不会有寄生二极管。因为在混合电压 I/O 电路中压焊点与 VDD 之间不能放置二极管，作用在 I/O 压焊点上的 PS 模式的 ESD 电压不能从压焊点被转移到 VDD 电源线，并且不能通过附加的电源轨（VDD 到 VSS）ESD 钳位电路泄放。这种压焊点上的 PS 模式的 ESD 电流完全通过快速反向击穿的层叠 NMOS 泄放。在 PS 模式 ESD 应力条件下沿着混合电压输出缓冲器的 ESD 电流如图 5-18b 中虚线部分所示。层叠配置中的 NMOS 器件具有比单个 NMOS 更高的触发电压（V_{t1}）、更高的快速反向保持电压（V_{sb}）和更低的二次击穿电流[74]。因此，这种采用层叠 NMOS 结构的混合电压 I/O 电路的 ESD 级别一般比单个 NMOS I/O 电路的 ESD 级别低很多[74,75]。所以混合电压 I/O 电路在 PS 模式 ESD 应力条件下一般有最低的 ESD 级别（采用人体模式 ESD 测试时一般 <2kV）。由于在 I/O 压焊点与 VDD 电源线之间没有寄生二极管，在 ND 模式应力条件下，混合电压 I/O 电路的 ESD 级别也较低。因此，混合 I/O 电

路的 ESD 保护设计主要着重于改善正 ESD 应力条件下的 ESD 级别。

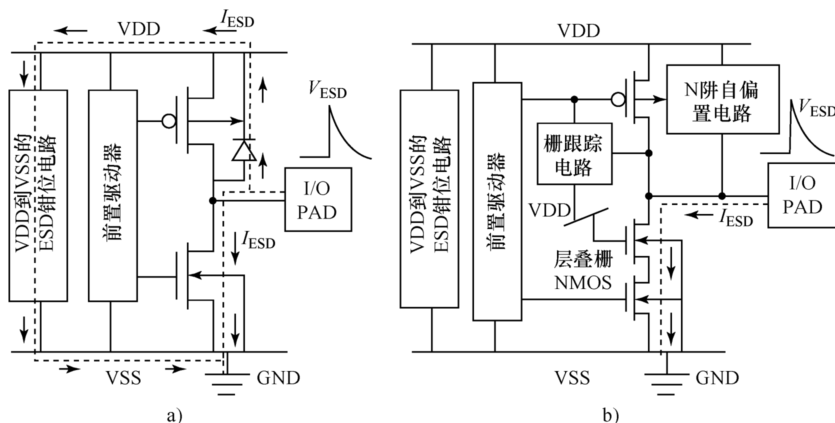


图 5-18　在 PS 模式 ESD 应力条件下缓冲器中 ESD 电流路径

a）传统的 CMOS 输出缓冲器　b）混合电压输出缓冲器

为了增加这种混合电压 I/O 电路的 ESD 级别，I/O 压焊点与 VDD 电源线之间采用了一些在层叠结构中添加多个二极管的设计[63,64]。然而，当混合电压 I/O 电路工作在高电压输入、高温环境时，必须采用附加电路设计，减小从压焊点通过层叠二极管到 VDD 的正向泄漏电流[76-79]。

5.5.3　混合电压 I/O 接口的 ESD 保护器件

本节将介绍一种新的 ESD 保护设计，通过采用层叠 NMOS 来触发 SCR 器件可以极大地改善混合电压 I/O 缓冲器的 ESD 鲁棒性[80,81]。针对混合电压 I/O 电路，在新的 ESD 保护电路中，SCR 器件采用层叠 NMOS 结构栅耦合电路技术，它可以与普通的 CMOS 工艺充分地兼容而不会引起栅氧化层可靠性问题。不需要采用厚的栅氧，针对 3V/5V　容限的混合电压 I/O 缓冲器的新 ESD 保护设计已经在 0.35μmCMOS 工艺中成功得到验证，并且应该在更小尺寸的工艺中也是适用的。

层叠 NMOS 触发晶闸管（SNTSCR）的剖面图与相应的版图分别如图 5-19a 与图 5-19b 所示。这种 SNTSCR 器件结构可以在典型的 CMOS 工艺中实现而不需要额外的工艺调整。在键合压焊点安排 SNTSCR 器件可以保护混合电压 I/O 电路免受 ESD 损伤。这种 SNTSCR 器件对应的等效电路如图 5-19c 所示。

在 SNTSCR 器件中，两个 NMOS 晶体管（Mn1 和 Mn2）以串接方式层叠在一起，其中 Mn1 的漏跨接在 N 阱与衬底之间的结上。在 I/O 压焊点与 VSS 之间由 P+扩散、N 阱、P 衬底和 n+扩散形成的一个横向 SCR 器件如图 5-19a 中虚

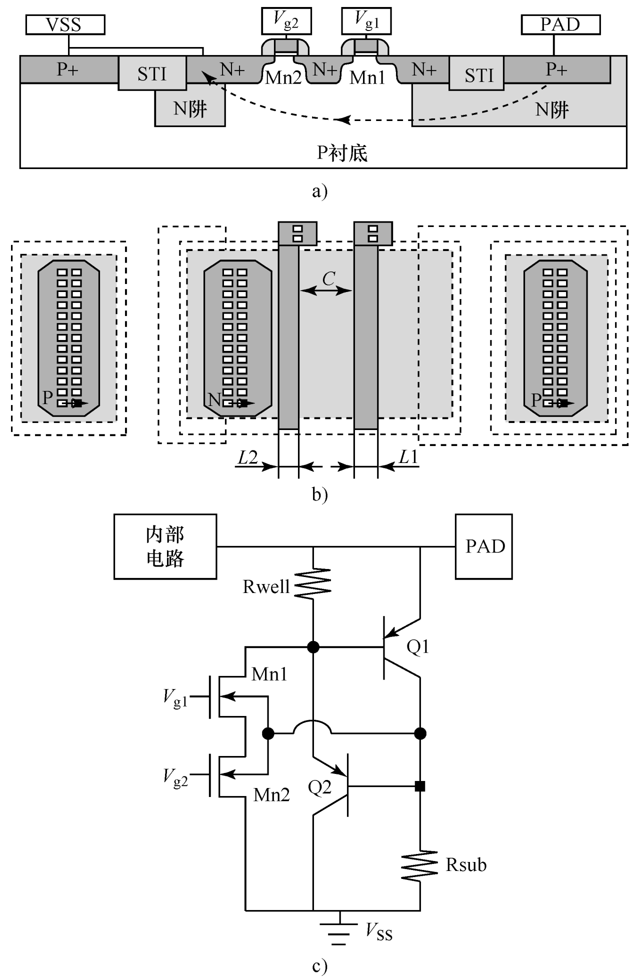


图 5-19　提出的采用 p 衬底 CMOS 工艺的 SNTSCR 器件

a）剖面图　b）相应的版图　c）等效电路

线所示。以层叠方式连接 Mn1 和 Mn2 的目的是为了在电路正常工作条件下能承受高电压水平的输入信号而不引起 SNTSCR 器件栅氧化层的可靠性问题。如果只采用单个 NMOS，例如像传统的 LVTSCR 那样[82]，当高压信号进入 I/O 压焊点时，栅氧化层上的电压将高于 VDD。如果传统的 LVTSCR 长期工作在这种混合电压 I/O 电路中，将引起栅氧化层可靠性问题。在 ESD 应力作用期间，栅偏置设计适当的 Mn1 和 Mn2 都将开启，触发横向 SCR 开启以泄放 ESD 电流。在 CMOS 工艺中无需厚栅氧器件，采用提出的 SNTSCR 器件作为保护混合 I/O 电路，不会引起栅氧化层可靠性问题。这种 ESD 结构对深亚微米工艺来说变得日益重要，其中薄栅氧器件被用在 I/O 结构中以减小高速应用时的引脚电容，并且

仍然与高压 IC 接口。

为了研究 SNTSCR 器件的特性，可以调整三个版图尺寸参数（图 5-19b 中所示的 C、$L1$ 和 $L2$）。其中 $L1$ 和 $L2$ 分别是 Mn1 和 Mn2 的沟道长度，C 是跨越在 Mn1 和 Mn2 器件共用的 n + 扩散区两侧多晶到多晶之间的间距。采用 0.35μm 工艺制造了版图参数不同但是沟道宽度固定为 60μm 的这种 SNTSCR 器件。其中 $C=0.5\mu m$ 而 $L1=L2=0.35\mu m$ 的一个 SNTSCR 器件在不同栅偏置 V_{g1} 和 V_{g2} 时的 I-V 特性曲线如图 5-20 所示。当栅偏置从 $V_{g1}=V_{g2}=0V$ 增加到 $V_{g1}=V_{g2}=0.5V$ 时，SNTSCR 器件的触发电压（V_t）从 10V 下降到 6V。当 $V_{g1}=V_{g2}>0.6V$ 时，Mn1 和 Mn2 都导通并触发 SNTSCR 开启，因此 V_t 下降到 1V ~ 2V。给 Mn1 和 Mn2 加适当的栅偏置，SNTSCR 器件的触发电压将低于混合电压 I/O 缓冲器中层叠 NMOS 的快速反向击穿电压。因此，新提出的采用 SNTSCR 器件的 ESD 保护电路能有效保护混合电压 I/O 缓冲器。

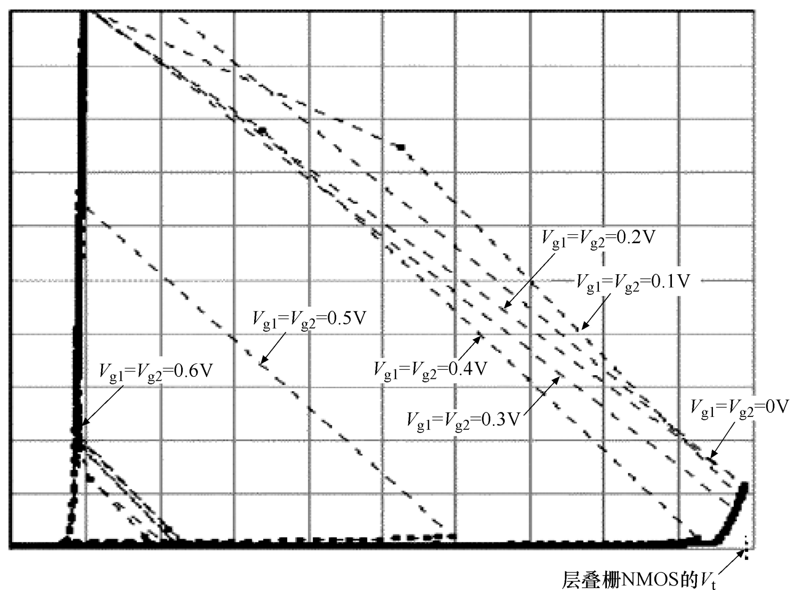


图 5-20　不同栅偏置时测得的 $C=0.5\mu m$ 而 $L1=L2=0.35\mu m$ 的一个实际 SNTSCR 器件的 $I-V$ 曲线（x 轴：1V/div；y 轴：1mA/div）

版图参数和栅偏置对 SNTSCR 器件在 PS 模式的 HBM ESD 级别的影响如图 5-21a和图 5-21b 所示。图 5-21a 显示的是固定沟道长度 $L1=L2=0.35\mu m$，SNTSCR 器件的 HBM ESD 级别与版图参数 C 之间的关系。图 5-21b 显示的是固定版图参数 $C=0.5\mu m$，SNTSCR 器件的 HBM ESD 级别与沟道长度 $L1$ 和 $L2$ 之间的关系。失效判据确定为偏置电压为 5V 时，泄漏电流大于 1μA。当版图参数 C、$L1$

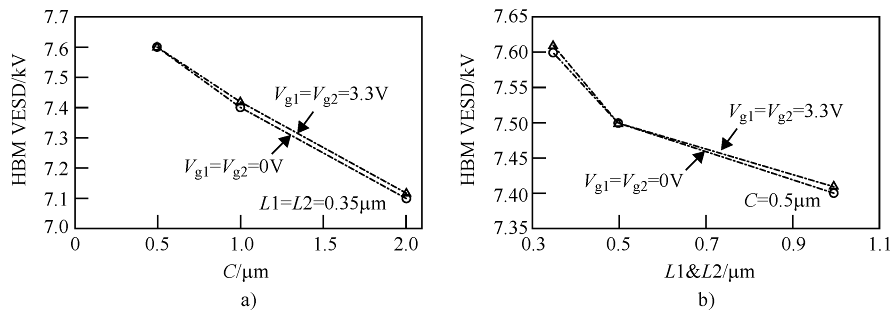


图 5-21　a）不同栅偏置（$V_{g1} = V_{g2}$）时 HBM ESD 级别与 SNTSCR 器件版图参数 C 之间的关系。失效判据：$V_{bias} = 5V$ 时，$I_{leakage} > 1\mu A$　b）不同栅偏置（$V_{g1} = V_{g2}$）时 HBM ESD 级别与 SNTSCR 器件版图参数 $L1$ 和 $L2$ 之间的关系。失效判据：$V_{bias} = 5V$ 时，$I_{leakage} > 1\mu A$

和 $L2$ 增加时，SNTSCR 器件的 HBM ESD 鲁棒性稍有衰退。版图参数（C、$L1$ 和 $L2$）较小的 SNTSCR 器件具有较窄的阳极到阴极间距，这意味着更小的导通电阻。因此，版图参数（C、$L1$ 和 $L2$）较小的 SNTSCR 器件有较高的 ESD 级别。栅偏置 V_{g1} 与 V_{g2} 并没有明显地改善 SNTSCR 器件的 ESD 级别（见图 5-21），但是它能较早的触发 SNTSCR 以泄放 ESD 电流。这样，栅偏置设置合适的 SNTSCR 器件可以有效保护 CMOS IC 的混合电压 I/O 缓冲器。

5.5.4　混合电压 I/O 接口的 ESD 保护电路设计

基于对 SNTSCR 器件的实验研究，采用 SNTSCR 的保护混合信号电压 I/O 缓冲器的 ESD 保护设计如图 5-22a 所示。其中设计了一个 ESD 检测电路可以在 ESD 应力条件下提供合适的栅偏置从而触发 SNTSCR 器件。另外，这个 ESD 检测电路必须使 SNTSCR 器件在 IC 处于正常工作条件时保持关断。使用栅耦合技术设计并考虑了栅氧化层可靠性问题的 ESD 检测电路如图 5-22b 所示。

在电路正常工作条件下，SNTSCR 处于关断状态，因此它不会干扰 I/O 压焊点的电压信号。在这种状态下，图 5-22b 所示的 Mn3 起电阻作用，将 Mn1 的栅压（V_{g1}）偏置到 VDD，而 Mn2 的栅通过电阻 R2 接地。当 I/O 压焊点加的是 5V 高输入电压时，Mn1 与 Mn2 晶体管之间共用的 n^+ 区的电压约为 VDD $-$ V_{thn}（V_{thn} 是 NMOS 器件的阈值电压）。所以 ESD 保护电路中的所有器件在电路正常工作期间都能满足不会使栅氧化层中电场太强的要求。当 I/O 压焊点的电压从 0V 变到 5V 时，通过电容 C1 耦合的电压也会使 Mn1 的栅压增加。因此图 5-22b 所示的 PMOS 器件（Mp）被设计为一旦 Mn1 的栅压增加到 VDD $+$ V_{thp}（V_{thp} 是 Mp 的阈值电压）时超过的电压便受到钳位。适当设计 ESD 检测电路可以确保 SNTSCR 器件在电路正常工作条件下处于关断状态。另外，即使 I/O 压焊点上有高输入电

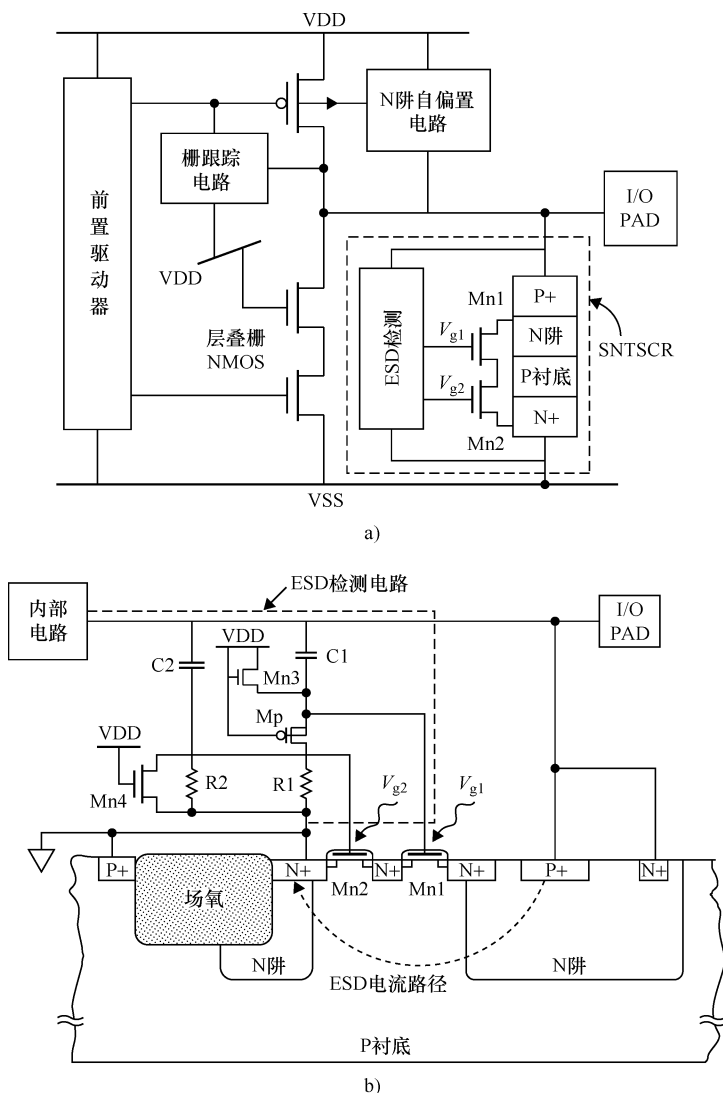


图 5-22　a）采用 SNTSCR 器件保护混合电压 I/O 缓冲器的 ESD 保护电路　b）保护混合电压 I/O 缓冲器的 ESD 检测电路设计中，SNTSCR 器件采用有较大的维持电阻和耦合电容设计范围

压，PMOS（Mp）可以进一步钳位 Mn1 的栅压以保证 Mn1 栅氧的可靠性。

电容（C1 和 C2）、电阻（R1 和 R2）、Mp 和 Mn3 组成一个给 SNTSCR 器件提供合适栅偏置的 ESD 检测电路。电容 C1（C2）被设计来耦合 ESD 瞬态电压到 Mn1（Mn2）的栅上以降低 SNTSCR 器件的触发电压。为了触发 SNTSCR 器件进入保持区，电阻 R1（R2）被设计为可以较长时间维持 Mn1（Mn2）栅上的耦

合电压。为了改善设计范围从而可以在常规的 CMOS 工艺中轻易地选择合适的维持电阻和耦合电容，在 Mn2 的栅和 VSS 之间的维持电阻 R2 上跨接了器件 Mn4。器件 Mn4 的栅被偏置到 VDD，但是考虑到与工艺相关的天线效应问题，最好通过扩散电阻将栅连接到 VDD 电源线。在电路正常工作条件下，Mn4 总是处于开启状态从而把耦合电压 V_{g2} 钳位到阈值电压（V_{thn}）之下，并且保持 Mn2 关断。所以，在电路正常工作条件下，可以保证 SNTSCR 被关断。

在 PS 模式 ESD 应力条件下，加到 I/O 压焊点的是一个较高的正 ESD 电压，VSS 接地而 VDD 悬空。在这种应力条件下，因为悬空的 VDD 电源线上的初始电压为零，使得 Mp 的栅被接地。因此 Mp 开启，而 Mn3 关断。电容 $C1$ 和 $C2$ 的作用是分别将 ESD 瞬态电压从 I/O 压焊点耦合到 Mn1 和 Mn2 的栅极。设计应该使得耦合电压高于阈值电压，从而保证混合电压 I/O 电路中的器件在被 ESD 能量损伤前能开启 Mn1 和 Mn2，将 SNTSCR 器件触发开启。另外，在 ESD 应力条件下，悬空的 VDD 电源线上的初始电平为零，因此 Mn4 关断。耦合到 Mn2 栅上的电压由维持电阻（R2）和耦合电容（C2）决定。当 SNTSCR 被触发开启时，ESD 电流主要通过 SNTSCR 器件从 I/O 压焊点泄放到 VSS。SNTSCR 器件具有较低的触发电压和较低的保持电压，可以安全的保护混合电压 I/O 电路的薄栅氧，还可以以较小的芯片面积达到较高的 ESD 级别。

ESD 检测电路的作用是在电路正常工作条件与 ESD 应力条件下给 SNTSCR 器件提供一个合适的栅偏置。为了得到合适的栅偏置，耦合电容（C1 和 C2）和维持电阻（R1 和 R2）值的确定非常重要。基于前面的工作原理，为了在各种 CMOS 工艺下都能保证电路如期望的那样工作，可以用 HSPICE 仿真方法精细调整确定合适的 C1、C2、R1 和 R2 值。

5.5.5 ESD 鲁棒性

通过测量，比较了采用和不采用 ESD 保护电路的混合电压 I/O 缓冲器在 PS 模式下的 HBM ESD 级别，结果如图 5-23a 所示。作为对比，还测试了具有不同层叠 NMOS 沟道宽度的混合电压 I/O 缓冲器。通过使用提出的采用 SNTSCR 器件的 ESD 保护电路可以使混合电压 I/O 缓冲器的 HBM ESD 级别（层叠 NMOS 沟道宽度为 120μm）获得明显的改善，从初始时的约 2kV 提升到高于 8kV。在图 5-23a中，由提出的 ESD 保护电路保护的所有混合电压 I/O 缓冲器中的 SNTSCR 器件宽度均为 60μm。

通过测量，比较了采用和不采用 ESD 保护电路的混合电压 I/O 缓冲器在 PS 模式下的机器模型（MM）的 ESD 级别，结果如图 5-23b 所示。作为对比，也测试了具有不同层叠 NMOS 沟道宽度的混合电压 I/O 缓冲器。结果表明，采用沟道宽度为 120μm 层叠 NMOS 的混合电压 I/O 缓冲器的 MM ESD 级别可以得到显著

改善。采用宽度仅为 60μm SNTSCR 的 ESD 保护电路，其 ESD 级别从约 200V 提升到约 800V。

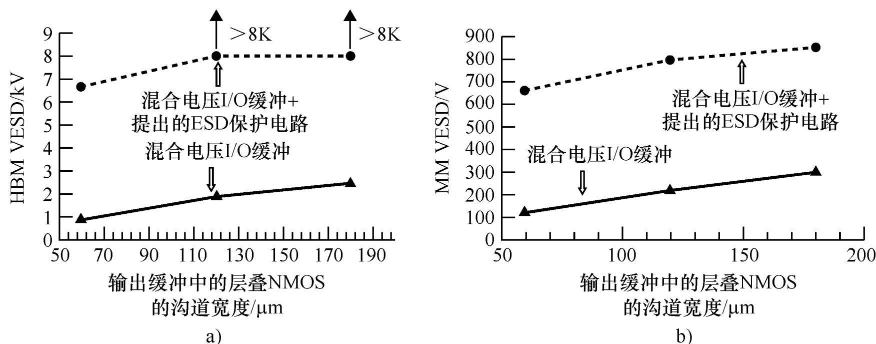
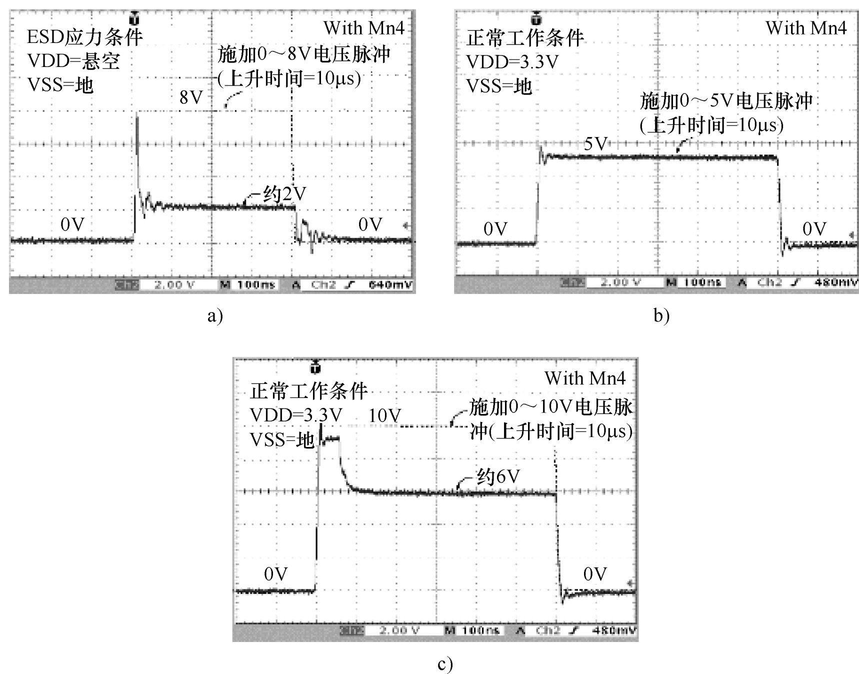


图 5-23　a）在混合电压 I/O 缓冲器中层叠 NMOS 的沟道宽度不同时，采用和不采用 ESD 保护电路的混合电压 I/O 缓冲器在 PS 模式下的 HBM ESD 鲁棒性比较　b）在混合电压 I/O 缓冲器中层叠 NMOS 的沟道宽度不同时，采用和不采用 ESD 保护电路的混合电压 I/O 缓冲器在 PS 模式下的 MM ESD 鲁棒性的比较

5.5.6　开启验证

为了验证混合电压 I/O 的 ESD 保护电路开启的有效性，当 VSS 相对接地而 VDD 悬空（为了仿真 PS 模式的 ESD 应力条件）时，给 I/O 压焊点加一个上升时间为 10ns 的 0V 到 8V 的尖锐上升电压脉冲。输出缓冲器中层叠 NMOS 的快速反向击穿电压约为 10V。这样一个加到 I/O 压焊点的 0V 到 8V 的脉冲并没有使混合电压输出缓冲器的层叠 NMOS 击穿。但是 0V 到 8V 的脉冲能触发开启 ESD 保护电路，从而得到一个如图 5-24a 所示的衰退的电压波形，其中 0V 到 8V 的脉冲被 SNTSCR 器件钳位到约 2V。图 5-24a 中从 8V 到 2V 约 10ns 的转换时间就是用 0.35μmCMOS 工艺实现的 SNTSCR 器件相应的开启速度。

在 VDD（VSS）为 3.3V（0V）的电路正常工作条件情况下，给 I/O 压焊点加一个上升时间为 10ns 的 0V 到 5V 的输入电压脉冲，采用数字示波器监测 I/O 压焊点上电压，结果如图 5-24b 所示，所加的 0V 到 5V 的波形并没有衰减。在电路正常工作期间，Mn4 的栅被偏置到 VDD（3.3V）。通过电容 C2 耦合的 Mn2 的栅电压通过开启的 Mn4 泄放到地，因此 SNTSCR 器件并没有被正常的 5V 输入信号触发开启。如果所加的电压脉冲进一步增加到 10V，在电路正常工作条件下，I/O 压焊点上测量的电压波形如图 5-24c 所示，所加的 0V 到 10V 的脉冲被钳位到约 6V，但是没有达到图 5-24a 所示的约 2V 的水平。在 0.35μmCMOS 工艺下，层叠 NMOS 的快速反向击穿电压约为 10V，而快速反向保持电压约为 6V。因此，图 5-24c 中所示的约 6V 的衰退电压是被混合电压 I/O 缓冲器中层叠



a)　　b)

c)

图 5-24　a）在 PS 模式 ESD 应力条件下，由上升时间为 10ns 的 0V 到 8V 电压脉冲触发，在 I/O 压焊点测量的电压波形（y 轴：2V/div；x 轴：100ns/div）　b）在正常工作条件下，VDD = 3.3V，VSS = 0V，由上升时间为 10ns 的低电平为 0V 到高电平为 5V 的电压波形　c）低电平为 0V 到高电平 10V 电压脉冲触发，在 I/O 压焊点测量的电压波形（y 轴：2V/div；x 轴：100ns/div）

NMOS 钳位在快速反向区的结果。如果 ESD 保护电路中的 SNTSCR 器件被触发开启，这个电压水平将被钳位到它的保持电压约 2V 左右。这个结果进一步证明了 ESD 保护电路中的器件 Mn4 可以被安全的加到 SNTSCR 器件中以保护混合电压 I/O 缓冲器而不会在电路正常工作条件下出乎意料地触发。

5.6　用于 ESD 保护的 SCR 器件

由于晶闸管（SCR）器件的保持电压 V_h 较低（在常规的 CMOS 工艺中 V_h 约为 1.5V），在 ESD 应力期间 SCR 器件的功耗（power ≈ $I_{ESD} \times V_h$）远小于其他 ESD 保护器件（如二极管、MOS、BJT 或场氧器件）的功耗。在 CMOS 集成电路中 SCR 器件能够以较小的版图面积维持较高的 ESD 级别，因此，长期以来被用于保护集成电路内部电路，使其免受 ESD 损伤。但是在小于 0.25 微米工艺中的 SCR 器件仍然有较高的开关电压（即触发电压，约为 22V），它一般大于输入级的栅氧击穿电压。此外，栅氧厚度、击穿时间或击穿电荷（QBD）将随着缩小

的 CMOS 工艺而减小。这种趋势迫切要求减小 SCR 的开关电压、提高它们的开启速度，从而可以有效的保护超薄栅氧免受潜在的损伤或破裂，特别是对快速器件充电模型（CDM）的 ESD 作用，成为在深亚微米中极难满足的一个规范。

本节将概述基于 SCR 器件的片上 ESD 保护。另外，还讨论在最大电源电压高于 1.5V 的 CMOS IC 产品中，避免基于 SCR 器件瞬态感应闩锁问题的解决方案。然而，在 IC 产品的最大电源电压小于 SCR 器件的保持电压时，这个闩锁问题就不存在了。例如，保持电压约为 1.6V 的单个 SCR 在最大电源电压为 1.2V 的 0.13μmCMOS 工艺中就能安全地被用于 ESD 保护而没有闩锁的危险。

5.6.1　SCR 器件的开启机制

SCR 器件等效电路图如 5-25a 所示。SCR 器件包含了一个横向 NPN 和一个纵向 PNP 双极晶体管，形成了一个两端、四层 PNPN（P+/N 阱/P 阱/N+）结构，这是 CMOS 工艺所固有的。SCR 器件的开关电压由 N 阱/P 阱结的雪崩击穿电压决定，其在 0.25μm CMOS 工艺中约为 22V，在 0.13μm CMOS 工艺中约为 18V。例如，当加到 SCR 阳极的正脉冲大于击穿电压，并且 SCR 的阴极接地，SCR 内将通过雪崩击穿机制产生电子和空穴电流。空穴电流将从 P 阱流到接地的 P+扩散，而电子电流将从 N 阱流到与 SCR 阳极连接的 N+扩散。只要 P 阱电阻（R_{pwell}）［N 阱电阻（R_{nwell}）］上的电压大于 0.7V，其中的 NPN（PNP）晶体管将开启，从而注入电子（空穴）电流，进一步偏置 PNP（NPN）晶体管，开始 SCR 闩锁动作，最后，成功触发 SCR 开启进入到闩锁状态从而通过正反馈再生机制泄放 ESD 电流[94,95]。

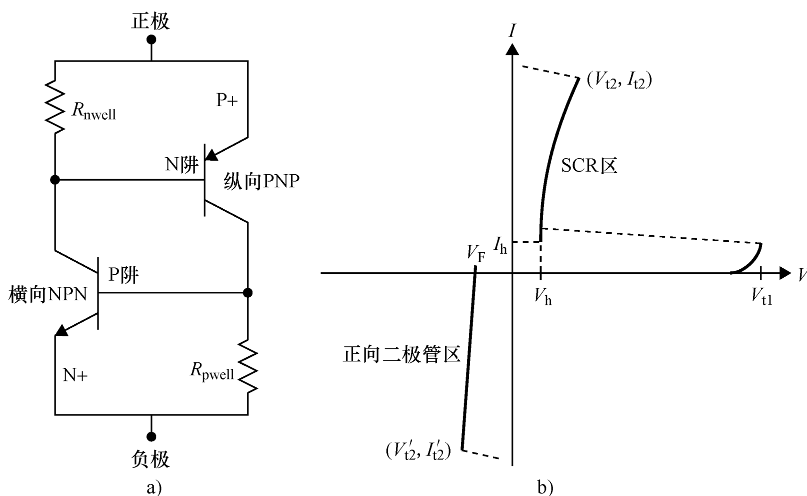


图 5-25　a）SCR 器件的等效电路图　b）CMOS 工艺下的 SCR 器件在正、负电压偏置下的 $I-V$ 特性

SCR 器件的直流 $I-V$ 特性如图 5-25b 所示。一旦 SCR 触发开启，保持 NPN 和 PNP 晶体管开启所要求的保持电流将通过闩锁的正反馈再生机制产生，而不包括雪崩击穿机制。所以 SCR 的保持电压（V_h）可以被减小到较低的数值，一般约 1.5V。当一个负电压被加到 SCR 结构的阳极一端时，SCR 结构内部的寄生二极管（N 阱/P 阱结）将被正向偏置，从而将负电压钳位到较低的约 1V（二极管的开启电压）。不管 ESD 能量是正还是负，SCR 器件都可以将 ESD 过应力钳位到较低的电压水平，所以在 CMOS IC 中，SCR 器件可以采用较小的面积维持最高的 ESD 鲁棒性。

5.6.2　基于 SCR 的 CMOS 片上 ESD 保护器件

低压触发 SCR[82,85]　为了更有效地保护输入甚至是输出级，开发出了一种低压触发 SCR（LVTSCR）。LVTSCR 的器件结构如图 5-26a 所示，0.25μm 工艺下 LVTSCR 的 $I-V$ 特性如图 5-26b 所示。采用 LVTSCR 器件作 ESD 保护电路的实例如图 5-26c 所示。在某些应用中，LVTSCR 的 N 阱被连接到输入压焊点。

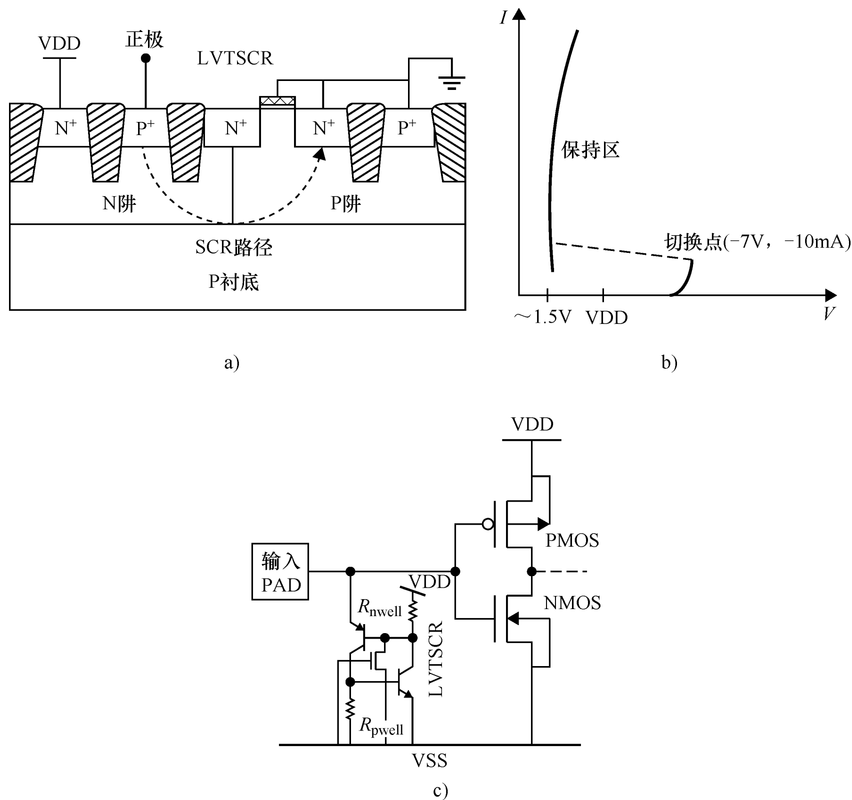


图 5-26　a）CMOS 工艺实现的低压触发 SCR（LVTSCR）器件结构　b）0.25μmCMOS 工艺 LVTSCR 的 $I-V$ 特性　c）采用 LVTSCR 器件的输入 ESD 保护电路

LVTSCR 的开关电压（约 7V）对应于插入到 LSCR 结构中的短沟 NMOS 器件的漏击穿或穿通电压，而不是 LSCR 器件的开关电压（22V）。有这么低开关电压的 LVTSCR 就能给 CMOS IC 输入或输出级提供有效的 ESD 保护，而不需要第二级 ESD 保护电路。因此使用 LVTSCR 的 ESD 保护电路的总版图面积将会大大节省。此外，为了能给 CMOS IC 输入或输出级的 PMOS 和 NMOS 都提供保护，已发明了一个互补的 LVTSCR 结构[99]以提供更好的 ESD 保护。

栅耦合 LVTSCR[86]　为了有效保护深亚微米 CMOS 工艺中的超薄栅氧，栅耦合技术被用来进一步减小 LVTSCR 的开关电压，而不涉及雪崩击穿机制。采用互补栅耦合 LVTSCR 器件［NMOS 触发 LSCR（NTLSCR）和 PMOS 触发 LSCR（PTLSCR）］的输入或输出压焊点的 ESD 保护电路如图 5-27a 所示。互补栅耦合 LVTSCR 的器件结构如图 5-27b 所示，采用 0.25μm 工艺的栅耦合 LVTSCR 的 $I-V$ 特性如图 5-27c 所示。SCR 器件的开关电压与 NTLSCR 器件的栅偏置电压之间的关系如图 5-27d 所示。图 5-27c 中的电容（C_n 和 C_p）值必须设计合理，以保证在电路正常工作条件下耦合的电压小于 NMOS/PMOS 的阈值电压，而在受到

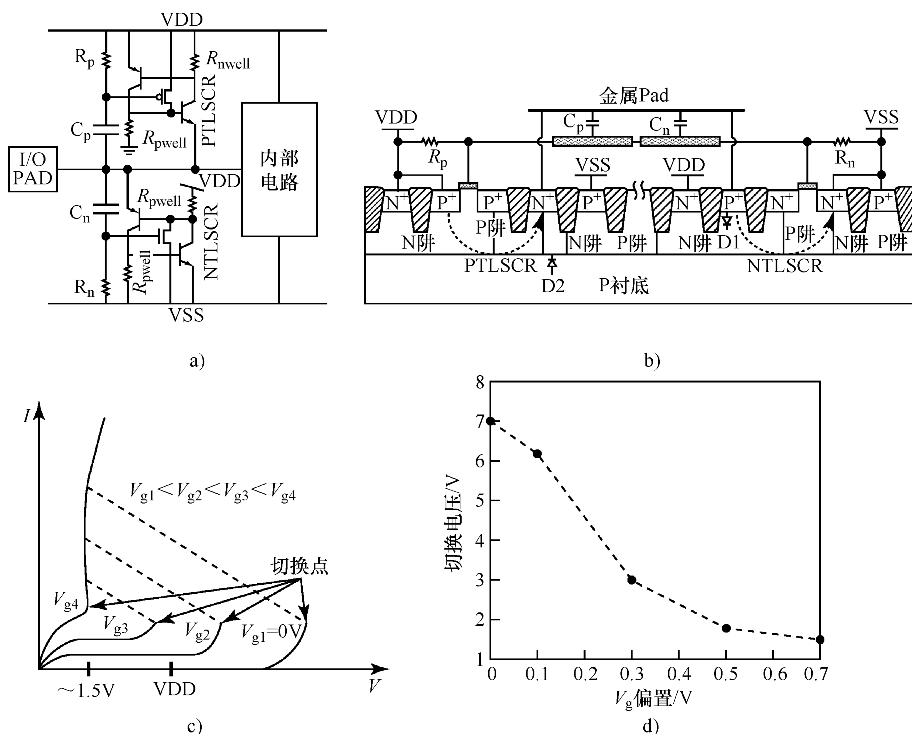


图 5-27　a）采用栅耦合 LVTSCR 的 ESD 保护电路　b）CMOS 工艺实现的栅耦合 LVTSCR 器件结构　c）0.25μm 工艺栅耦合 LVTSCR 的 $I-V$ 特性　d）SCR 器件开关电压与 NTLSCR 器件栅偏置电压之间的关系

ESD 作用的条件下耦合电压则应该大于 NMOS/PMOS 的阈值电压[20]。栅耦合 NTLSCR（PTLSCR）的开关电压可以通过 SCR 器件结构中的短沟 NMOS（PMOS）栅上的耦合电压来调整。在 LVTSCR 中，短沟 NMOS/PMOS 栅上较高的耦合电压会导致 LVTSCR 较低的开关电压。因此，栅耦合 LVTSCR 器件能快速的泄放 ESD 电流以更有效地保护输入或输出级的超薄栅氧。

栅接地 NMOS 触发 SCR[88,89] 采用栅接地 NMOS 触发的 SCR 是片上 ESD 保护电路的另一个选择。一个类似于 GGNMOS 结构的 NMOS 晶体管被用作一个外部触发器件以触发开启 GGSCR。与 LVTSCR 相比，GGSCR 中的外部触发 NMOS 的漏极被直接耦合到压焊点，而它的栅和源被耦合到 P 衬底（NPN 的基极)。一个使用 GGSCR 器件作为输入 ESD 保护电路的实例如图 5-28a 所示。图 5-28b 所示为 GGSCR 版图的俯视图[105]。当 ESD 电压被加到图 5-28a 所示的 I/O 压焊点时，外部触发 NMOS 将进入雪崩击穿状态并开始将触发电流注入到 P 衬底和多晶电阻。只要 NPN 的基极电压大于 0.7V，GGSCR 将触发开启。图 5-28b 中多晶电阻用来控制触发电流和保持电流，并阻止 GGSCR 的误触发。实验结果表明，与阳极到阴极间距较大的 LVTSCR 相比，阳极到阴极间距较小的 GGSCR 具有较低的保持电压、较高的 It2、更好的 dV/dt 触发能力和较快的开启速度。

衬底触发的 SCR[90,91] SCR 器件的开启机制本质上是电流触发事件。当电流被加到 SCR 器件的基极或衬底时，它能被快速地触发进入闩锁状态。有报道说已经开发出了使用衬底触发技术的 P 型衬底触发的 SCR（P_STSCR）和 N 型衬底触发的 SCR（N_STSCR）器件实现 ESD 保护。P_STSCR 和 N_STSCR 的器件结构分别如图 5-29a 和图 5-29b 所示。与传统的横向 SCR 器件结构相比，这种结构在 P_STSCR 器件结构的 P 阱中插入了一个附加的 P^+ 扩散，并且将其接出作为 P_STSCR 器件的 P 触发器节点。对 N_STSCR，在 N_STSCR 器件结构的 N 阱中插入了一个附加的 N^+ 扩散，并且将其接出作为 N_STSCR 器件的 N 触发器节点。P_STSCR 和 N_STSCR 的 $I-V$ 特性分别如图 5-29c 和图 5-29d 所示。随着衬底/阱触发电流的增加，P_STSCR/N_STSCR 器件的开关电压可以被减小到它的保持电压。在 0.25μm 工艺中，STSCR 的开启时间在上升时间为 10ns 的 5V 电压脉冲下会减小到约 10ns。开关电压较低的 STSCR 器件可以更快地把 ESD 电压钳位到较低的电压水平，以充分保护输入级的超薄栅氧免受 ESD 过应力的损伤。使用 P_STSCR和 N_STSCR 器件的输入或输出压焊点的 ESD 保护电路如图 5-29e 所示。在有采用 VDD 和 VSS 电源的电路在正常工作条件下，inv_1（inv_2）的输入被偏置到 VDD（VSS）。因此，由于 inv_1（inv_2）中 NMOS（PMOS）的开启，不管输入信号是逻辑高电平还是逻辑低电平，inv_1（inv_2）的输出都被偏置到 VSS（VDD）。P_STSCR（N_STSCR）的 p 触发器（n 触发器）节点被 inv_1(inv_2)的输出偏置到 VSS（VDD）。所以，在电路正常工作情况下，P_STSCR 和

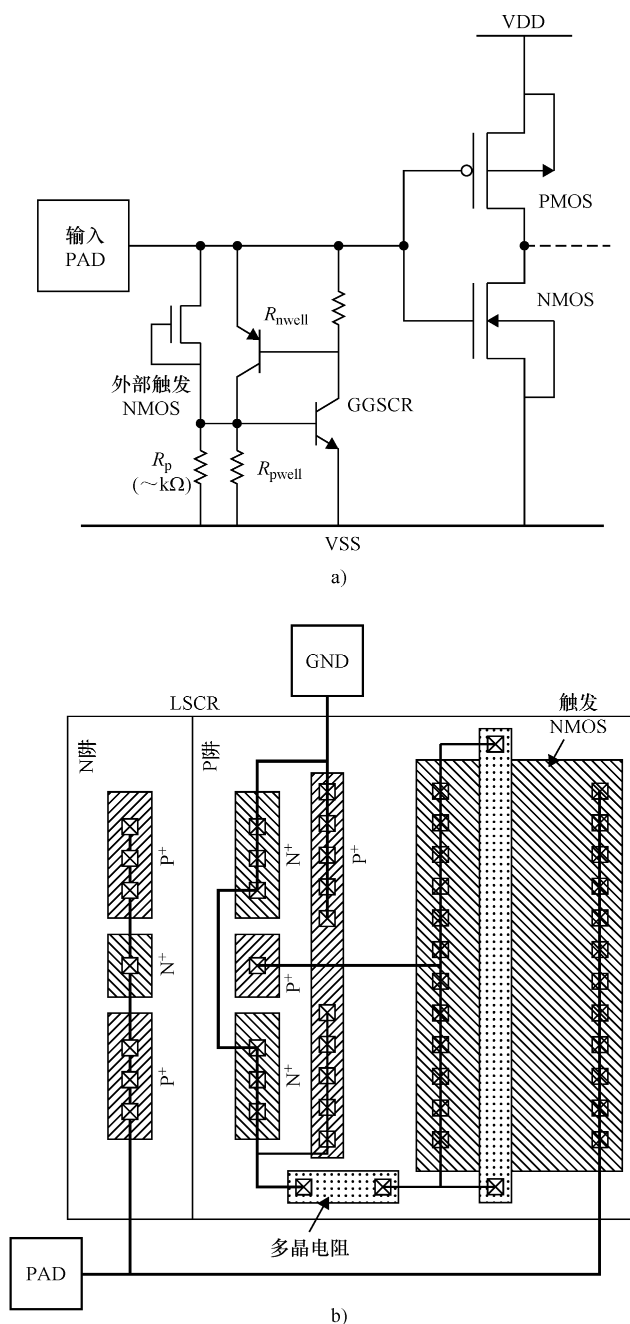


图 5-28　a）采用 GGSCR 器件的输入 ESD 保护电路
b）CMOS 工艺实现的 GGSCR 版图俯视图

N_STSCR 器件可以保证处于关断状态。在图 5-29e 所示的正脉冲加于 VSS（PS）的 ESD 作用条件下（VSS 接地而 VDD 悬空），inv_1 的输入最初处于零电平悬空状态；因此，由于压焊点上的正 ESD 电压作用，inv_1 的 PMOS 将导通。所以 inv_1 的输出被 ESD 能量充电，产生流向 P_STSCR 器件 P 触发器节点的触发电流。因此，P_STSCR 器件被触发开启，ESD 电流通过 P_STSCR 器件从I/O压焊点泄放到

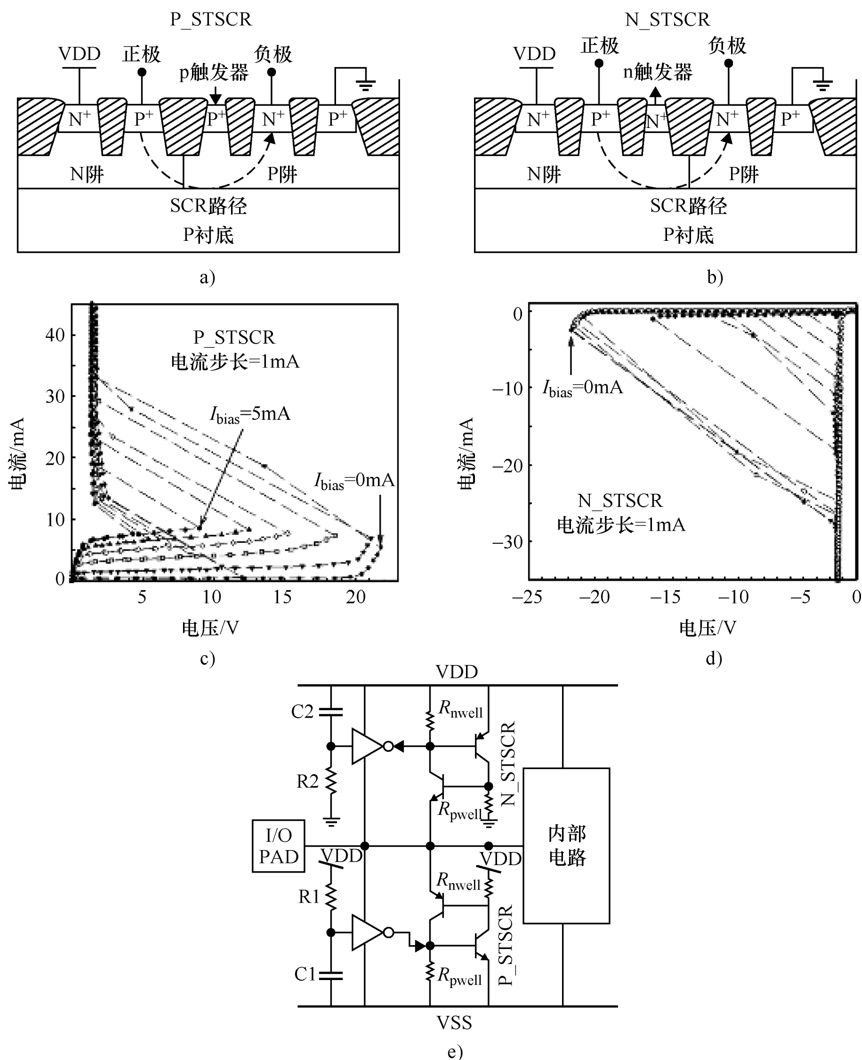


图 5-29　a）P 型衬底触发的 SCR（P_STSCR）器件结构　b）N 型衬底触发的 SCR（N_STSCR）器件结构　c）采用 0.25μmCMOS 工艺的 P_STSCR 的 $I—V$ 特性　d）采用 0.25μmCMOS 工艺的 N_STSCR 的 $I-V$ 特性　e）采用 P_STSCR 和 N_STSCR 实现的 I/O 压焊点 ESD 保护电路

接地的 VSS 引脚。在负脉冲加于 VSS（ND）模式下的 ESD 作用情况下（VDD 接地而 VSS 悬空），N_STSCR 的工作原理类似。另外，还发明了一种采用哑元栅结构的 STSCR 器件[108,109]可以进一步减小开关电压并且提高 STSCR 的开启速度。采用哑元栅结构的 STSCR 的双极电流增益大于采用浅沟槽隔离技术（STI）结构的 STSCR 的双极电流增益，因此采用哑元栅结构的 STSCR 的触发效率优于采用 STI 的 STSCR。

双触发 SCR[94]　进一步减小 LSCR 器件开关电压、更有效提高 LSCR 开启速度的另一种方式是双触发技术。双触发 SCR（Double-Triggered SCR，DTSCR）的器件结构如图 5-30a 所示。在 DTSCR 器件结构的 P 阱和 N 阱中又添加了 P^+ 和 N^+ 扩散，并且被接出作为 DTSCR 器件的 P 触发器和 N 触发器的节点。图 5-30b 汇总了在不同的 N 阱触发电流下，DTSCR 器件的开关电压与衬底触发电流之间的关系。如果衬底和 N 阱触发电流被分别同步加到 p 触发器和 n 触发器的节点，DTSCR 器件的开关电压能被有效减小到较低的电压水平。采用 DTSCR 器件实现的 I/O ESD 保护电路如图 5-30c 所示。基于 RC 延迟原理，在 PS（ND）ESD 作

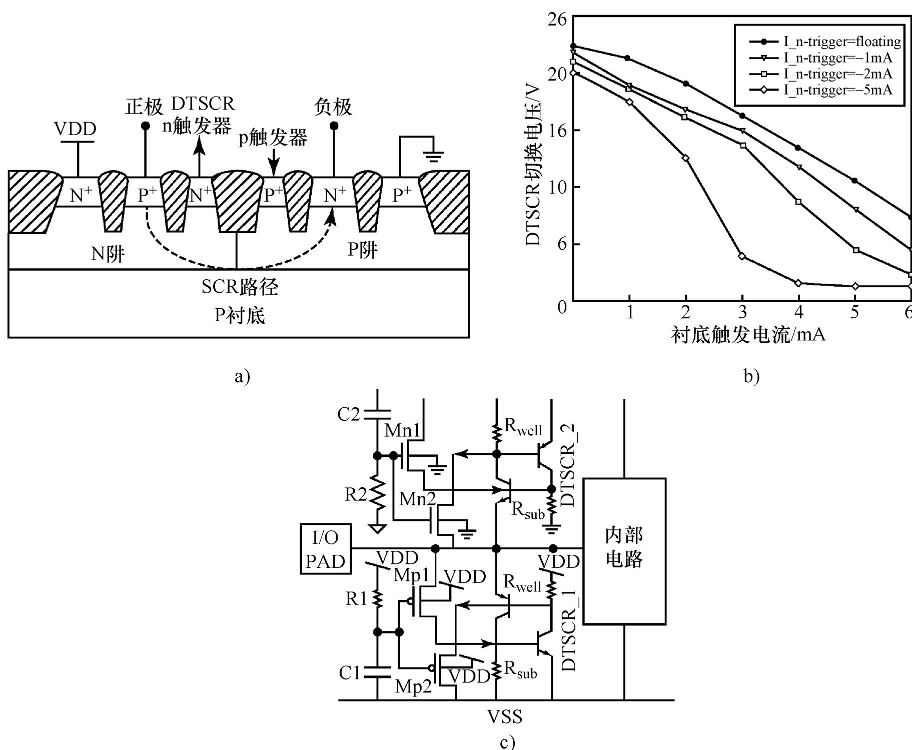


图 5-30　a）双触发 SCR（DTSCR）的器件结构　b）在不同 N 阱触发电流下，采用 0.25μmCMOS 工艺的 DTSCR 器件维持电压与衬底触发电流之间的关系　c）采用 DTSCR 器件的 I/O 压焊点 ESD 保护电路

用条件下，衬底和 N 阱触发电流可以分别由 Mp1 和 Mp2（Mn1 和 Mn2）产生。因此，图 5-30c 所示的双触发电流的 DTSCR 可以被更快的触发以泄放 ESD 电流。在采用 VDD 和 VSS 电源的电路正常工作条件下，Mp1 和 Mp2（Mn1 和 Mn2）的栅通过电阻 R1（R2）偏置到 VDD（VSS）。因此，无论输入信号为高电平或低电平，Mp1、Mp2、Mn1 和 Mn2 都处于关断状态。DTSCR 器件的 p 触发器（n 触发器）节点通过寄生电阻 R_{sub}（R_{well}）保持在 VSS（VDD），所以在电路正常工作条件下，可以确保这种 DTSCR 处于关断状态。根据 0.25μmCMOS 工艺的试验结果，当一个 0V 到 5V 的电压脉冲被加到 DTSCR 的阳极时，如果 DTSCR 的 p 触发器节点上有 1.5V 的正电压脉冲，它的开启时间约为 37ns。但是当一个 -5V 的负电压脉冲被加到它的 n 触发器的节点时，DTSCR 的开启时间可以进一步减小到约 12ns。SCR 器件中被用来阻挡 STI 的假栅结构可以被应用到 DTSCR 结构中进一步减小开关电压、更加有效地提高 DTSCR 器件的开启速度。

本征 NMOS 触发 SCR[95] 本征 NMOS 是直接在亚 0.25μmCMOS 工艺中的 p 型轻掺杂衬底上形成，而通常的 NMOS（PMOS）则是在采用 p 型衬底的双阱工艺中的重掺杂 P 阱（N 阱）中形成。本征 NMOS 和横向 SCR 可以被结合在一起形成一种新的 ESD 保护器件，称为本征 NMOS 触发 SCR（Native-NMOS-triggered SCR，NANSCR）。这种器件的优势是具有更低的开关电压和更快的开启速度。NANSCR 的器件结构如图 5-31a 所示。本征 NMOS 的栅被连接到了负偏压电路（Negative Bias Circuit，NBC）[96]，使其在电路正常工作条件下关断 NANSCR。NANSCR 与 LVTSCR 的 DC $I-V$ 曲线对比如图 5-31b 所示。在 0.13μm 金属硅化物 CMOS 工艺下，沟道宽度相同、沟道长度为 0.3μm 的 NANSCR 的开关电压（约 4V）小于沟道长度为 0.13μm 的 LVTSCR 的开关电压（约 5V）。采用衬底触发技术，NANSCR 的开关电压可以随着本征 NMOS 的 W/L 比的增加而进一步下降。采用 NANSCR 器件的输入或输出压焊点的 ESD 保护电路如图 5-31c 所示。在电路正常工作条件下，所有 NANSCR 器件中的本征 NMOS 的栅被同一个 NBC 偏置，使 NANSCR 器件关断。所以 NANSCR 器件 NANSCR_1 和 NANSCR_2 将不会干扰 I/O 电路的功能。在 PS 模式 ESD 作用条件下，NANSCR_1 中的本征 NMOS 栅悬空，已经开启的本征 NMOS 产生的衬底触发电流快速将 NANSCR_1 触发开启。所以正 ESD 电流可以从 I/O 压焊点通过 NANSCR_1 快速泄放到接地的 VSS。根据 0.13μm 工艺的实验结果，NANSCR 的开启速度快于 LVTSCR。而且，NANSCR 承受的 CDM ESD 级别（5V/μm^2）高于 LVTSCR（2.33V/μm^2），因此它更适合用于保护纳米技术中的超薄栅氧。

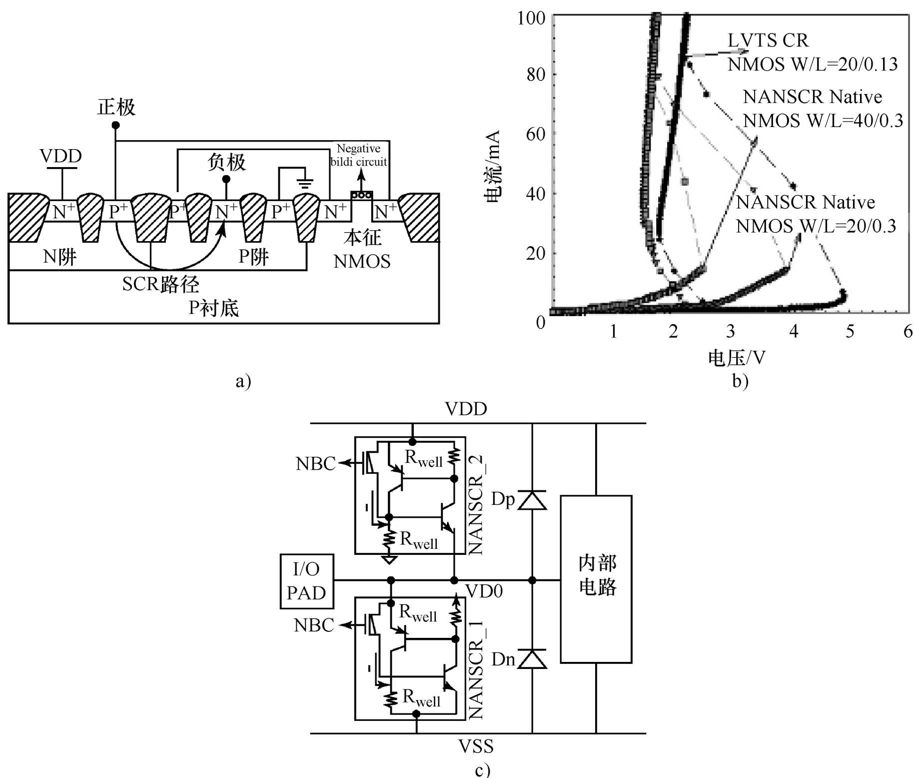


图 5-31　a）本征 NMOS 触发 SCR（NANSCR）的器件结构　b）0. 13μm 工艺实现的 NANSCR 与 LVTSCR 的 $I-V$ 特性比较　c）采用 NANSCR 器件的 ESD 保护电路

5. 6. 3　SCR 闩锁工程

为了使基于 SCR 的器件提供有效 ESD 保护，需要有足够低的开关电压，但是还应该避免瞬态感应闩锁问题[87]。在 CMOS IC 处于电路正常工作条件下，有两种方法可以避免基于 SCR 的开启电压较低的器件被噪声电压误开启。图 5-32a 所示为通过增加基于 SCR 的低压触发器件的触发电流来避免闩锁，但是其开关电流和保持电流不变。由于具有较高的触发电流，基于 SCR 的低触发电压器件，如 LVTSCR，对压焊点上的过冲或下冲噪声脉冲就具有足够的噪声裕量。采用 0. 6μm 工艺给 LVTSCR 结构增加了一个旁路二极管，已成功开发出高电流 NMOS 触发横向 SCR（HINTSCR，High-current NMOS-triggered lateral SCR）[99]器件，将触发电流增加到 218. 5mA。这种 HINTSCR 在 3V 应用时有大于 VDD + 12V 的噪声裕量。另外，还报导有将 GGSCR 改进为具有高保持电流的 SCR（HHI – SCR）器件[89]，在 0. 1μm CMOS 工艺中，通过将 GGSCR 中几千欧姆的外部多晶电阻调整到 HHI-SCR 中的约仅为 10Ω，使其保持电流达到约 70mA。

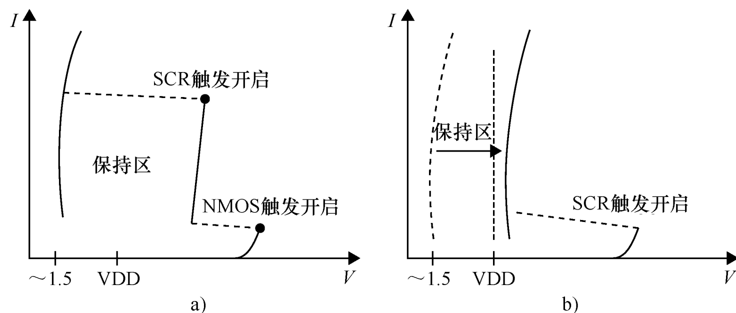


图 5-32　采用基于 SCR 器件的 ESD 保护设计中的两种克服闩锁问题的方案

a）增加触发电流　b）增加保持电压以避免基于 SCR 的器件被噪声脉冲误触发

避免闩锁的另一种方法是增加基于 SCR 器件的保持电压，使其高于最大 VDD 电源电压，如图 5-32b 所示。通过使用外延衬底，SCR 器件的保持电压可以增加到避免闩锁问题[116]。但是 CMOS 晶圆的制造费用将增加。通过将某些元件（如二极管或者 SCR 器件）串联，其总的串联电压降可以使得体 CMOS 工艺中基于 SCR 器件的总保持电压得到提升。通过设计适当的辅助触发，开关电压和电流可以被保持在较低电压水平。在 0.35μm 金属硅化物 CMOS 技术下，设计了一种级联 LVTSCR[117] 结构，增加了保持电压（ > VDD），而不会导致其 ESD 鲁棒性的衰退。另外，据报道，在 0.25μm 工艺技术下采用层叠 STSCR 器件设计或采用层叠 STSCR 器件与层叠二极管链设计的 ESD 保护电路具有 7kV 的 HBM ESD 级别，并且不存在闩锁问题。最近，对单个 SCR 器件，针对 ESD 保护（低保持电压）条件和电路正常工作（较高的保持电压）条件动态调整保持电压[102]。据报道，采用动态保持电压 SCR（Dynamic holding voltage SCR，DHVSCR）作 ESD 保护器件时，具有较高的闩锁免疫力。DHVSCR 器件的结构如图 5-33a 所示。在 DHVSCR 器件结构中而不是 LSCR 结构中插入了一个 PMOS 和一个 NMOS。0.25μm CMOS 工艺中 DHVSCR 在电路正常工作条件下和 ESD 放电条件下的 $I-V$ 特性如图 5-33b 所示。PMOS 和 NMOS 的栅电压（V_{g1} 和 V_{g2}）在电路正常工作条件下，被偏置到 2.5V（VDD），但是在 ESD 作用时则偏置到 0V。电路正常工作条件下 DHVSCR 的保持电压和保持电流分别为 2.8V 和 172mA。因此，电路正常工作条件下的 DHVSCR 将不会处于闩锁状态。然而，ESD 作用时 DHVSCR 的保持电压和保持电流分别降为 202V 和 91mA，因此 DHVSCR 也能将 ESD 过应力钳位到较低的电压水平，维持较高的 ESD 级别。通过控制与 SCR 结构结合在一起的 PMOS 和 NMOS 的栅压可以调整 DHVSCR 的保持电压和保持电流。

然而，随着 CMOS 按比例缩小技术的发展，按照“电场不变”的要求缩放器件尺寸，CMOS IC 的电源电压也随之缩减，同时也减小了功耗。对采用 0.13μm 金

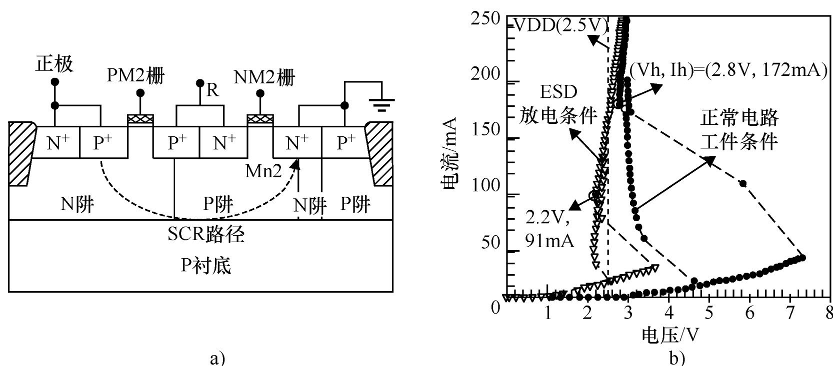


图 5-33　a）CMOS 工艺实现的动态保持电压 SCR（DHVSCR）的器件结构　b）采用 0.25μmCMOS 工艺的 DHVSCR 在电路正常工作条件下和 ESD 作用条件下的 $I-V$ 特性

属硅化物工艺实现的 CMOS IC，内部电路的最大电源电压已减小到了 1.2V，因此基于 SCR 的器件本身所固有的闩锁问题就不存在了。与其他 ESD 保护器件相比，采用基于 SCR 的开关电压较低的器件具有最高的 ESD 鲁棒性、最小的版图面积以及不存在闩锁问题，所以是片上 ESD 保护的最佳选择。然而，在这种超薄栅氧的 0.13μm 的 CMOS 工艺中，为了保护这种更薄的栅氧，应该提高 SCR 器件的开启速度以快速泄放 ESD 过应力电压。对于纳米 CMOS 工艺中超薄栅氧，采用 NAN-SCR[95] 和哑元栅结构改善基于 SCR 的器件的开启速度[92] 是较好选择。

必须更好地设计基于 SCR 器件的开启电压和开启速度，以充分有效地保护输入级中的超薄栅氧，特别是抗拒快速 CDM ESD 作用。在最大电源电压低于 1.2V 的纳米 CMOS 工艺中，基于 SCR 的器件的开关电压和开启速度将是采用基于 SCR 器件的片上 ESD 保护电路整体性能的决定性因素。对纳米 CMOS 工艺，为了保护 I/O 电路的超薄栅氧，要求基于 SCR 的器件必须有足够低的开关电压和足够快的开启速度。

5.7　总结

随着 CMOS 工艺不断缩小到纳米尺寸，一个芯片中，例如片上系统（SoC）中集成有更多的电路和功能。在尺寸缩小的 CMOS 工艺中，栅氧变得更薄，这使其更易受 ESD 过应力破坏。SoC 把更多的电路和功能集成到一个芯片中，其面积一般更大，这就导致 SoC 整个芯片中一般具有更大的体电容存储 CDM 电荷。另外，SoC 一般有几百个引脚，这使每个 I/O 单元的面积受到极大的限制，随之就限制了这种高引脚数 SoC 中 ESD 保护器件的版图面积。SoC 经常有多个不同的电源引脚，这会引起了电路模块之间接口电路发生不希望的 ESD 失效。因此，当工艺尺寸缩

小进入深亚微米时，ESD 问题将变得更加严重，特别是对 CDM ESD 放电。

为了有效保护采用纳米工艺的 SoC，必须提高 ESD 保护器件的开启速度，以便在 ESD 应力损伤内部电路前快速泄放 ESD 电流。如果片上 ESD 钳位器件（NMOS）具有负的阈值电压（或者近似为 0V），当 IC 受到 ESD 脉冲作用时，ESD 钳位器件将处于“已开启”状态以泄放 ESD 电流。但是在电路正常工作时，可以使用一个附加的负偏压来关断 ESD 钳位器件。采用“已开启”特性，纳米 CMOS 工艺的超薄栅氧仍然可以被安全地保护。

对采用多个不同电源引脚的 SoC，一个能避免 SoC 中内部电路或接口电路失效的全片 ESD 保护方案非常重要。电源线的布局安排对 SoC 的 ESD 级别有着很大的影响。应该采用合适的 ESD 连接单元来连接芯片上不同的电源线。在 ESD 保护器件中，就承受 ESD 应力而言，SCR 具有最高的面积效率。但是四层结构的 SCR 的开启速度一般较低，如果开启速度太慢将不能有效保护超薄栅氧的 SoC。如果采用经过改进了的设计或触发电路，使 SCR 器件有足够快的开启速度，它将是 VDD 电压低于 1.2V 情况下 SoC 中 ESD 保护电路的最好选择，而且不会出现闩锁问题。

参考文献

[1] *ESD Association Standard Test Method for Electrostatic Discharge Sensitivity Testing: Human Body Model-Component Level*, ESD STM 5.1, ESD Association, 2001. Rome, NY

[2] *ESD Association Standard Test Method for Electrostatic Discharge Sensitivity Testing: Machine Model-Component Level*, ESD STM 5.2, ESD Association, 1999. Rome, NY

[3] *ESD Association Standard Test Method for Electrostatic Discharge Sensitivity Testing: Charged Device Model-Component Level*, ESD STM 5.3.1, ESD Association, 1999. Rome, NY

[4] C. Duvvury, R. N. Rountree, and O. Adams, “Internal chip ESD phenomena beyond the protection circuit,” *IEEE Trans. Electron Devices*, Vol. 35, pp. 2133–2139, 1988.

[5] M.-D. Ker, Whole-chip ESD protection design with efficient VDD-to-VSS ESD clamp circuit for submicron CMOS VLSI, *IEEE Trans. Electron Devices*, Vol. 46, No. 1, pp. 173–183, Jan. 1999.

[6] M.-D. Ker and H.-H. Chang, ESD bus lines in CMOS ICs for whole-chip ESD protection, U.S. patent 6,144,542, Nov. 2000.

[7] M.-D. Ker, ESD protection circuit for mixed mode integrated circuits with separated power pins, U.S. patent 6,075,686, June 2000.

[8] M.-D. Ker and H.-H. Chang, ESD protection scheme for mixed-voltage CMOS integrated circuits, U.S. patent 6,002,568, Dec. 1999.

[9] M.-D. Ker and H.-H. Chang, Whole-chip ESD protection for CMOS ICs using bi-directional SCRs, U.S. patent 6,011,681, Jan. 2000.

[10] C. Jiang, E. Nowak, and M. Manley, Process and design for ESD robustness in deep submicron CMOS technology, *Proceedings of the IEEE International Reliability Physics Symposium*, pp. 233–236, 1996.

[11] C. Duvvury, R. N. Rountree, and R. A. McPhee, ESD protection: design and layout issues for VLSI circuits, *IEEE Trans. Ind. Appl.*, Vol. 25, pp. 41–47, 1989.

[12] A. Stricker, D. Gloor, and W. Fichtner, Layout optimization of an ESD-protection n-MOSFET by simulation and measurement, *Proceedings of the EOS/ESD Symposium*, pp. 205–211, 1995.

[13] M.-D. Ker, C.-Y. Wu, and T.-S. Wu, Area-efficient layout design for CMOS output transistors, *IEEE Trans. Electron Devices*, Vol. 44, No. 4, pp. 635–645, Apr. 1997.

[14] M.-D. Ker, T.-Y. Chen, and H.-H. Chang, New layout design for submicron CMOS output transistors to improve driving capability and ESD robustness, *J. Microelectron. Reliab.*, Vol. 39, No. 3, pp. 415–424, June 1999.

[15] M. Mergens, K. Verhaege, C. Russ, J. Armer, P. Jozwiak, G. Kolluri, and L. Avery, Multi-finger turn-on circuits and design techniques for enhanced ESD performance and width-scaling, *Proceedings of the EOS/ESD Symposium*, pp. 1–11, 2001.

[16] T.-Y. Chen and M.-D. Ker, Analysis on the dependence of layout parameters on ESD robustness of CMOS devices for manufacturing in deep-submicron CMOS process, *IEEE Trans. Semicond. Manuf.*, Vol. 16, No. 3, pp. 486–500, Aug. 2003.

[17] C. Jiang, E. Nowak, and M. Manley, Process and design for ESD robustness in deep submicron CMOS technology, *Proceedings of the IEEE International Reliability Physics Symposium*, pp. 233–236, 1996.

[18] C. Duvvury, C. Diaz, and T. Haddock, Achieving uniform nMOS device power distribution for sub-micron ESD reliability, *Technical Digest, IEEE International Electron Devices Meeting*, pp. 131–134, 1992.

[19] C. Duvvury and C. Diaz, Dynamic gate coupling of NMOS for efficient output ESD protection, *Proceedings of the IEEE International Reliability Physics Symposium*, pp. 141–150, 1992.

[20] M.-D. Ker, C.-Y. Wu, T. Cheng, and H.-H. Chang, Capacitor-couple ESD protection circuit for deep-submicron low-voltage CMOS ASIC, *IEEE Trans. VLSI Syst.*, Vol. 4, No. 3, pp. 307–321, 1996.

[21] H.-H. Chang and M.-D. Ker, Improved output ESD protection by dynamic gate floating design, *IEEE Trans. Electron Devices*, Vol. 45, No. 9, pp. 2076–2078, Sept. 1998.

[22] J. Chen, A. Amerasekera, and C. Duvvury, Design methodology and optimization of gate-driven NMOS ESD protection circuits in submicron CMOS processes, *IEEE Trans. Electron Devices*, Vol. 45, No. 12, pp. 2448–2456, 1998.

[23] M.-D. Ker, T.-Y. Chen, and C.-Y. Wu, Design of cost-efficient ESD clamp circuits for the power rails of CMOS ASICs with substrate-triggering technique, *Proceedings of the IEEE International ASIC Conference and Exhibit*, pp. 287–290, 1997.

[24] M.-D. Ker, T.-Y. Chen, C.-Y. Wu, H. Tang, K.-C. Su, and S.-W. Sun, Novel input ESD protection circuit with substrate-triggering technique in a 0.25-μm shallow-trench-isolation CMOS technology, *Proceedings of the IEEE International Symposium on Circuits and Systems*, Vol. 2, pp. 212–215, 1998.

[25] C. Duvvury, S. Ramaswamy, V. Gupta, A. Amerasekera, and R. Cline, Substrate pump NMOS for ESD protection applications, *Proceedings of the EOS/ESD Symposium*, pp. 7–17, 2000.

[26] M.-D. Ker, T.-Y. Chen, and C.-Y. Wu, Substrate-triggered ESD clamp devices for using in power-rail ESD clamp circuits, *Solid-State Electron.*, Vol. 46, No. 5, pp. 721–734, Apr. 2002.

[27] M.-D. Ker and T.-Y. Chen, Substrate-triggered ESD protection circuit without extra process modification, *IEEE J. Solid-State Circuits*, Vol. 38, No. 2, pp. 295–302, Feb. 2003.

[28] M.-D. Ker and T.-Y. Chen, Substrate-triggered technique for on-chip ESD protection design in a 0.18-μm salicided CMOS process, *IEEE Trans. Electron Devices*, Vol. 50, No. 4, pp. 1050–1057, Apr. 2003.

[29] T.-Y. Chen and M.-D. Ker, Investigation of the gate-driven effect and substrate-triggered effect on ESD robustness of CMOS devices, *IEEE Trans. Device Mater. Reliab.*, Vol. 1, No. 4, pp. 190–203, Dec. 2001.

[30] M.-D. Ker and H.-C. Jiang, Whole-chip ESD protection strategy for CMOS integrated circuits in nanotechnology, *Proceedings of the IEEE International Conference on Nanotechnology*, pp. 325–330, 2001.

[31] J.-S. Lee, Method for fabricating an electrostatic discharge protection circuit, U.S. patent 5,672,527, Sept. 1997.

[32] T.-Y. Huang, Method for making an integrated circuit structure, U.S. patent 5,529,941, June 1996.

[33] C.-C. Hsue and J. Ko, Method for ESD protection improvement, U.S. patent 5,374,565, Dec. 1994.

[34] R.-Y. Shiue, C.-S. Hou, Y.-H. Wu, and L.-J. Wu, ESD implantation scheme for 0.35 μm 3.3 V 70A gate oxide process, U.S. patent 5,953,601, Sept. 1999.

[35] T. Lowrey and R. Chance, Static discharge circuit having low breakdown voltage bipolar clamp, U.S. patent 5,581,104, Dec. 1996.

[36] J.-J. Yang, Electrostatic discharge protection circuit employing MOSFETs having double ESD implantations, U.S. patent 6,040,603, Mar. 2000.

[37] M.-D. Ker and C.-H. Chuang, ESD implantations in 0.18-μm salicided CMOS technology for on-chip ESD protection with layout consideration, *Proceedings of the International Symposium on the Physical and Failure Analysis of Integrated Circuits*, pp. 85–90, 2001.

[38] M.-D. Ker, T.-Y. Chen, and H.-H. Chang, ESD implantation method in deep-submicron CMOS technology for high-voltage-tolerant applications with light-doping concentrations, U.S. patent 6,514,839, Feb. 2003.

[39] M.-D. Ker, H.-C. Hsu, and J.-J. Peng, ESD implantation for sub-quarter-micron CMOS technology to enhance ESD robustness, *IEEE Trans. Electron Devices*, in press, Volume 50, Issue 10, Oct 2003, pg 2126–2134. Oct. 2003.

[40] K. Mistry, N-channel clamp for ESD protection in self-aligned silicided CMOS process, U.S. patent 5,262,344, Nov. 1993.

[41] H. Ooka, Semiconductor integrated circuit device including two types of MOSFETs having source/drain region different in sheet resistance from each other, U.S. patent 5,283,449, Feb. 1994.

[42] T. Randazzo and B. Larsen, Input/output transistors with optimized ESD protection, U.S. patent 5,493,142, Feb. 1996.

[43] J.-J. Wang, P.-C. Shieh, and P.-N. Tseng, Method of forming a resistor for ESD protection in self aligned silicide process, U.S. patent 5,547,881, Aug. 1996.

[44] J.-S. Lee, Method for fabricating an electrostatic discharge protection circuit, U.S. patent 5,672,527, Sept. 1997.

[45] M. Ma, ESD protection using selective siliciding techniques, U.S. patent 5,744,839, Apr. 1998.

[46] T.-Y. Chen, M.-D. Ker, and H.-H. Chang, Method of fabricating ESD protection device by using the same photolithographic mask for both the ESD implantation and the silicide blocking regions, U.S. patent 6,444,404, Sept. 3, 2002.

[47] D. Scott, P. Bosshart, and J. Gallia, Circuit to improve electrostatic discharge protection, U.S. patent 5,019,888, May 1991.

[48] K. Lee, A. Lee, M. Marmet, and K. Ouyang, Electro-static discharge protection circuit with bimodal resistance characteristics, U.S. patent 5,270,565, Dec. 1993.

[49] G.-L. Lin and M.-D. Ker, Fabrication of ESD protection device using a gate as a silicide blocking mask for a drain region, U.S. patent 6,046,087, Apr. 2000.

[50] C.-S. Kim et al., A novel NMOS transistor for high performance ESD protection devices in 0.18 μm CMOS technology utilizing salicide process, *Proceedings of the EOS/ESD Symposium*, pp. 407–412, 2000.

[51] A. Amerasekera and C. Duvvury, The impact of technology scaling on ESD robustness and protection circuit design, *Proceedings of the EOS/ESD Symposium*, pp. 237–245, 1994.

[52] S. Daniel and G. Krieger, Process and design optimization for advanced CMOS I/O ESD protection devices, *Proceedings of the EOS/ESD Symposium*, pp. 206–213, 1990.

[53] S. G. Beebe, Methodology for layout design and optimization of ESD protection transistors, *Proceedings of the EOS/ESD Symposium*, pp. 265–275, 1996.

[54] I. E. Opris, Bootstrapped pad protection structure, *IEEE J. Solid-State Circuits*, pp. 300–301, Feb. 1998.

[55] M.-D. Ker, T.-Y. Chen, C.-Y. Wu, and H.-H. Chang, ESD protection design on analog pin with very low input capacitance for high-frequency or current-mode applications, *IEEE J. Solid-State Circuits*, Vol. 35, No. 8, pp. 1194–1199, Aug. 2000.

[56] *Star–Hspice User's Manual: Applications and Examples*, Avanti, Campbell, CA 1998.

[57] M.-D. Ker, H.-C. Jiang, and C.-Y. Chang, Design on the low-capacitance bond pad for high-frequency I/O circuits in CMOS technology, *IEEE Trans. Electron Devices*, Vol. 48, No. 12, pp. 2953–2956, Dec. 2001.

[58] M.-D. Ker and C.-Y. Chang, ESD protection design for CMOS RF integrated circuits using polysilicon diodes, *J. Microelectron. Reliab.*, Vol. 42, No. 6, pp. 863–872, June 2002.

[59] M.-D. Ker and T.-Y. Chen, Layout design to minimize voltage-dependent variation on input capacitance of an analog ESD protection circuit, *J. Electrostat.*, Vol. 54, No. 1, pp. 73–93, Jan. 2002.

[60] T. Furukawa, D. Turner, S. Mittl, M. Maloney, R. Serafin, W. Clark, J. Bialas, L. Longenbach, and J. Howard, Accelerated gate-oxide breakdown in mixed-voltage I/O circuits, *Proceedings of the IEEE International Reliability Physics Symposium*, pp. 169–173, 1997.

[61] E. Takeda and N. Suzuki, An empirical model for device degradation due to hot-carrier injection, *IEEE Electron Device Lett.*, Vol. 4, pp. 111–113, 1983.

[62] S. H. Voldman, ESD protection in a mixed voltage interface and multi-rail disconnected power grid environment in 0.5- and 0.25-μm channel length CMOS technologies, *Proceedings of the EOS/ESD Symposium*, pp. 125–134, 1994.

[63] S. Dabral and T. J. Maloney, *Basic ESD and I/O Design*, Wiley, New York, 1998.

[64] M. Hargrove et al., High-performance sub-0.08 μm CMOS with dual gate oxide and 9.7 ps inverter delay, *Technical Digest, IEEE International Electron Devices Meeting*, pp. 22.4.1–22.4.4, 1998.

[65] S. Poon et al., A versatile 0.25-micron CMOS technology, *Technical Digest, IEEE International Electron Devices Meeting*, pp. 28.2.1–28.2.4, 1998.

[66] M. Takahash, T. Sakurai, K. Sawada, K. Nogami, M. Ichida, and K. Matsud, 3.3 V–5 V compatible I/O circuit without thick gate oxide, *Proceedings of the IEEE Custom Integrated Circuits Conference*, pp. 23.3.1–23.3.4, 1992.

[67] M. Pelgrom and E. Dijkmans, A 3/5 V compatible I/O buffer, *IEEE J. Solid-State Circuits*, Vol. 30, pp. 823–825, 1995.

[68] J. Conner, D. Evans, G. Braceras, J. Sousa, W. Abadeer, S. Hall, and M. Robillard, Dynamic dielectric protection for I/O circuits fabricated in a 2.5-V CMOS technology interfacing to a 3.3-V LVTTL bus, *Technical Digest, International Symposium on VLSI Circuits*, pp. 119–120, 1997.

[69] G. Singh and R. Salem, High-voltage-tolerant I/O buffers with low-voltage CMOS process, *IEEE J. Solid-State Circuits*, Vol. 34, pp. 1512–1525, 1999.

[70] H. Sanchez, J. Siegel, C. Nicoletta, J. Nissen, and J. Alvarez, A versatile 3.3/2.5/1.8-V CMOS I/O driver built in a 0.2-μm 3.5-nm Tox 1.8-V CMOS technology, *IEEE J. Solid-State Circuits*, Vol. 34, pp. 1501–1511, 1999.

[71] A.-J. Annema, G. Geelen, and P. de Jong, 5.5-V I/O in a 2.5-V 0.25-μm CMOS technology, *IEEE J. Solid-State Circuits*, Vol. 36, pp. 528–538, 2001.

[72] M.-D. Ker and C.-S. Tsai, Design of 2.5 V/5 V mixed-voltage CMOS I/O buffer with only thin oxide device and dynamic *n*-well bias circuit, *Proceedings of the 2003 IEEE International Symposium on Circuits and Systems*, Vol. 5, pp. 97–100, 2003.

[73] T. Maloney and W. Kan, Stacked PMOS clamps for high voltage power supply protection, *Proceedings of the EOS/ESD Symposium*, pp. 70–77, 1999.

[74] W. Anderson and D. Krakauer, ESD protection for mixed-voltage I/O using NMOS transistors stacked in a cascode configuration, *Proceedings of the EOS/ESD Symposium*, 1998, pp. 54–71.

[75] J. Miller, M. Khazhinsky, and J. Weldon, Engineering the cascoded NMOS output buffer for maximum V_{t1}, *Proceedings of the EOS/ESD Symposium*, pp. 308–317, 2000.

[76] T. J. Maloney and S. Dabral, Novel clamp circuits for IC power supply protection, *Proceedings of the EOS/ESD Symposium*, pp. 1–12, 1995.

[77] S. H. Voldman, G. Gerosa, V. Gross, N. Dickson, S. Furkay, and J. Slinkman, Analysis of snubber-clamped diode-string mixed voltage interface ESD protection network for advanced microprocessors, *Proceedings of the EOS/ESD Symposium*, pp. 43–61, 1995.

[78] T. J. Maloney, K. Parat, N. Clark, and A. Darwish, Protection of high voltage power and programming pins, *Proceedings of the EOS/ESD Symposium*, pp. 246–254, 1997.

[79] M.-D. Ker and W.-Y. Lo, Design on the low-leakage diode string for using in the power-rail ESD clamp circuits in a 0.35-μm silicide CMOS process, *IEEE J. Solid-State Circuits*, Vol. 35, pp. 601–611, 2000.

[80] M.-D. Ker and C.-H. Chuang, Stacked-NMOS triggered silicon-controlled rectifier for ESD protection in high/low-voltage-tolerant I/O interface, *IEEE Electron Device Lett.*, Vol. 23, No. 6, pp. 363–365, June 2002.

[81] M.-D. Ker and C.-H. Chuang, ESD protection design for mixed-voltage CMOS I/O buffers, *IEEE J. Solid-State Circuits*, Vol. 37, No. 8, pp. 1046–1055, Aug. 2002.

[82] A. Chatterjee and T. Polgreen, A low-voltage triggering SCR for on-chip ESD protection at output and input pads, *IEEE Electron Device Lett.*, Vol. 12, pp. 21–22, 1991.

[83] M.-D. Ker and C.-Y. Wu, Modeling the positive-feedback regenerative process of CMOS latch-up by a positive transient pole method: I. Theoretical derivation, *IEEE Trans. Electron Devices*, Vol. 42, pp. 1141–1148, 1995.

[84] M.-D. Ker and C.-Y. Wu, Modeling the positive-feedback regenerative process of CMOS latch-up by a positive transient pole method: II. Quantitative evaluation, *IEEE Trans. Electron Devices*, Vol. 42, pp. 1149–1155, 1995.

[85] M.-D. Ker, C.-Y. Wu, and H.-H. Chang, Complementary-LVTSCR ESD protection circuit for submicron CMOS VLSI/ULSI, *IEEE Trans. Electron Devices*, Vol. 43, pp. 588–598, 1996.

[86] M.-D. Ker, H.-H. Chang, and C.-Y. Wu, A gate-coupled PTLSCR/NTLSCR ESD protection circuit for deep-submicron low-voltage CMOS IC's, *IEEE J. Solid-State Circuits*, Vol. 32, pp. 38–51, 1997.

[87] G. Weiss and D. Young, Transient-induced latch-up testing of CMOS integrated circuits, *Proceedings of the EOS/ESD Symposium*, pp. 194–198, 1995.

[88] C. Russ, M. P. J. Mergens, J. Armer, P. Jozwiak, G. Kolluri, L. Avery, and K. Verhaege, GGSCR: GGNMOS triggered silicon controlled rectifiers for ESD protection in deep submicron CMOS processes, *Proceedings of the EOS/ESD Symposium*, pp. 22–31, 2001.

[89] M. P. J. Mergens, C. C. Russ, K. G. Verhaege, J. Armer, P. C. Jozwiak, and R. Mohn, High holding current SCRs (HHI-SCR) for ESD protection and latch-up immune IC operation, *Proceedings of the EOS/ESD Symposium*, pp. 10–17, 2002.

[90] M.-D. Ker and K.-C. Hsu, Substrate-triggered SCR device for on-chip ESD protection in fully silicided subquarter-micrometer CMOS process, *IEEE Trans. Electron Devices*, Vol. 50, pp. 397–405, 2003.

[91] M.-D. Ker and K.-C. Hsu, Latch-up-free ESD protection design with complementary substrate-triggered SCR devices, *IEEE J. Solid-State Circuits*, Vol. 38, No. 8, pp. 1380–1392, Aug. 2003.

[92] K.-C. Hsu and M.-D. Ker, Improvement on turn-on speed of substrate-triggered SCR device by using dummy-gate structure for on-chip ESD protection, *Proceedings of the International Conference on Solid State Devices and Materials*, pp 440–441, Sept. 16–18 2003.

[93] M.-D. Ker and G.-L. Lin, Low-voltage-triggered electrostatic discharge protection device and relevant circuitry, U.S. patent 6,465,848, Oct. 2002.

[94] M.-D. Ker and K.-C. Hsu, SCR device with double-triggered technique for on-chip ESD protection in sub-quarter-micron silicided CMOS process, *IEEE Trans. Device Mater. Reliab.*, in press, Volume 3, Issue 3, Sept. 2003, pg 58–68. Sept. 2003.

[95] M.-D. Ker and K.-C. Hsu, Native-NMOS-triggered SCR (NANSCR) for ESD protection in 0.13-μm CMOS integrated circuits, Proc. of 2004 IEEE International Reliability Physics Symposium (IRPS), Phoenix, Arizona, 2004, pp. 381–386, April 25–29.

[96] M.-D. Ker, C.-Y. Chang, and H.-C. Jiang, Design of negative charge pump circuit with polysilicon diodes in a 0.25-μm CMOS process, *Proceedings of the IEEE AP-ASIC Conference*, pp. 145–148, 2002.

[97] A. Z. H. Wang and C.-H. Tsay, An on-chip ESD protection circuit with low trigger voltage in BiCMOS technology, *IEEE J. Solid-State Circuits*, Vol. 36, pp. 40–45, 2001.

[98] M.-D. Ker and C.-Y. Wu, CMOS on-chip electrostatic discharge protection circuit using four-SCR structures with low ESD-trigger voltage, *Solid-State Electron.*, Vol. 37, pp. 17–26, 1994.

[99] M.-D. Ker, Lateral SCR devices with low-voltage high-current triggering characteristics for output ESD protection in submicron CMOS technology, *IEEE Trans. Electron Devices*, Vol. 45, pp. 849–860, 1998.

[100] G. Notermans, F. Kuper, and J.-M. Luchis, Using an SCR as ESD protection without latch-up danger, *Microelectron. Reliab.*, Vol. 37, pp. 1457–1460, 1997.

[101] M.-D. Ker and H.-H. Chang, How to safely apply the LVTSCR for CMOS whole-chip ESD protection without being accidentally triggered on, *Proceedings of the EOS/ESD Symposium*, pp. 72–85, 1998.

[102] Z.-P. Chen and M.-D. Ker, Dynamic holding voltage SCR (DHVSCR) device for ESD protection with high latch-up immunity, *Proceedings of the International Conference on Solid State Devices and Materials*, in press, pp. 160–161 Sept. 16–18 2003.

第 6 章　输入/输出设计

6.1　引言

在先进的工艺中，输入和输出缓冲器设计变得日益困难。在给定工艺下，确定如何设计 I/O 缓冲器时，设计者面临着大量的必须区分优先顺序的变量。这些必须考虑的问题包括：

（1）I/O 标准；

（2）采用不同电压水平的核心电路与 I/O 缓冲器之间的信号转换；

（3）静电放电（ESD）保护方案；

（4）性能要求；

（5）引脚电容［与上述问题（3）和（4）紧密相关］；

（6）电迁移；

（7）I/O 开关噪声；

（8）版图；

（9）端接；

（10）阻抗匹配；

（11）前置放大；

（12）均衡化。

对每个工艺节点，I/O 缓冲器的性能都必须相应增加。对比分析表明，核心芯片部分的性能提高明显高于 I/O 缓冲器，这也说明了 I/O 缓冲器性能的提升变得更加困难。这主要是因为标准 I/O 要求的电源电压一般比核心电路高得多。为此总是要求采用栅氧较厚的器件（或器件级联），这样，如果栅的尺寸相同，则厚栅氧器件的驱动能力就减小。为了保证能提供足够的驱动，要求采用面积更大的输出器件，满足信号边缘速率的要求。这样就要增大扩散掺杂面积，使引脚电容增加，最终使信号边缘速率下降。最后，将会达到一个饱和点，这时继续增加缓冲器尺寸将不能再提高信号边缘速率。下面几节将分析这方面的一些问题并介绍在更先进的工艺中 IC 设计者当前改善 I/O 性能的一些趋势。本章是以标准 CMOS 工艺为基础的。如果采用更特殊的工艺，如锗硅工艺，将会引起新的设计问题。

6.2　I/O 标准

在过去的 30 年产生了许多的 I/O 标准。主要的标准包括：先进图形端口（AGP）、电流型逻辑（CML）、发射极耦合逻辑（ECL）、正发射极耦合逻辑（PECL）、发射接收逻辑（GTL）、高速收发器逻辑（HSTL）、残余连续终结逻辑（SSTL）、低压 CMOS（LVCMOS）、低压差分信号（LVDS）、超高速连接（它是 LVDS 标准的子集）、低压发射极耦合逻辑（LVPECL）和低压晶体管晶体管逻辑（LVTTL）。这些标准中许多是单端的，使性能受到严重制约。目前真正的高速接口电路主要使用的 I/O 标准是 HSTL、SSTL、LVDS、LVPECL（以及其他形式的 ECL）和 CML。CML 缓冲器一般用于速度最高的应用中。

目前 I/O 缓冲器中使用的两个基本信号配制是：单端和差分。单端信号的工作速度不超过 600～800Mbit/s，主要是满足存储器接口电路的需求，虽然采用其他一些技术已经可以获得更高的速率[5]。对 LVDS 标准，差分标准具有 1～1.5Gbit/s的数据传输速率，并且在 CML 标准下速率可高达 10Gbit/s（并可能更高）。差分 I/O 标准可以分为两大类：并行总线和串行连接。LVDS 标准是一种比较典型的并行总线，而 CML 是串行连接。图 6-1 所示为一个典型的高速串行系统。采用 LVDS 标准将并行信号从左侧输入到芯片中。数据流在传送到串行端之前使用 8B/10B 标准进行编码。采用这种编码方案保证时钟恢复的 0→1 和 1→0 的转换，并且使得 0 和 1 的个数平均来说基本相等，保持直流平衡。尽管这种编码方案增加了一些带宽开销，但是它依旧是使用最普遍的互连编码形式。例如，如果使用 8B/10B 编码并且期望的数据速率达 1Gbit/s，为了解决与编码相关的开销问题，实际的传输速率必须为 1.25Gbit/s。

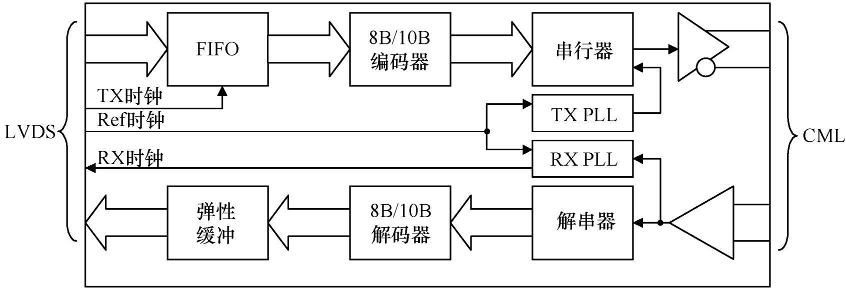


图 6-1　对并行数据采用 LVDS 信号而串行数据
采用 CML 信号的典型串行系统图

6.3　信号传输

设计时必须最先做出的决定之一是 I/O 器件使用的氧化层厚度。对于给定的工艺，有几个厚度供选用，但是从工艺成本和成品率的角度看，减少实际选用的氧化层厚度数目可以极大地减小产品的总成本。多数情况下，使用两种氧化层厚度就可以实现核心电路性能和 I/O 支持的电压水平之间的折衷。一般情况下，采用尽可能薄的氧化层，使较小的器件有更大的电流驱动能力，可以极大地减小引脚电容，并使 I/O 具有最好的性能。如果核心电路的工作电压与 I/O 工作电压不同，在两个电源域之间的转换就成为一个问题。这个问题已在第 4 章详细讨论过。在许多情况下，为了避免 I/O 噪声被耦合到内部电路，I/O 工作在与核心电路完全独立的电源和地配置下。图 6-2 所示为一个将数据从核心电路电平转换到输出缓冲电平的可能方案简图。这里，信号传递到噪声问题更严重的输出缓冲的地线之前，相对于核心电路噪声更小的电源，信号一直被作为参考，直到信号边缘速率增加。在两个地线之间转换时，地弹效应对信号抖动有着巨大的影响，而保持快速的边缘速率可以减轻这种效应。第二种方式是在 IO 地线（$V_{SS\ IO}$）与核心电路地线（$V_{SS\ core}$）之间形成局部短路，如图 6-2 中两个节点之间虚线所示。局部短路给电流提供了返回到驱动源的路径，能减小地弹效应。这种方式有一定的风险，因为一些不希望的电流会从 I/O 的地线流到核心电路的地线，在核心电路中形成了不希望的噪声。本节中没有涉及的关键概念是核心电压和 I/O 电压之间的转换并不是一个简单的问题，必须给予极大的关注。必须对转换级进行仿真以评价其整体性能。好的设计可以获得 20～100ps/V 的灵敏度因子。

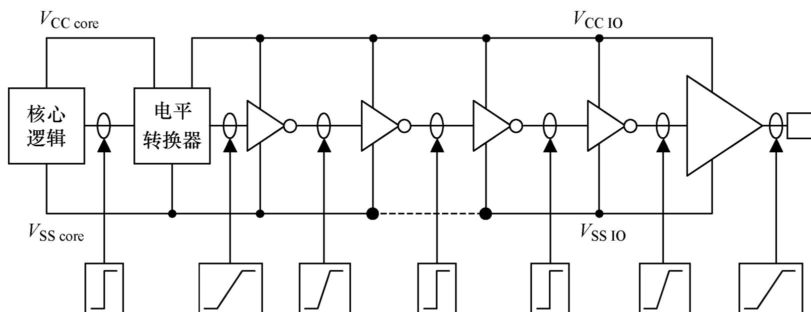


图 6-2　输出缓冲器信号传输方案

6.3.1　单端缓冲器

单端缓冲器的基本结构并没有随着工艺尺寸的缩小而变化。主要的变化是每

个工艺节点都必须支持的更高的信号速率。如果采用固定的接口，如 1.8V STTL，性能问题将会减少。这使得输出缓冲器可以采用更薄的栅氧器件进行设计，提供更好的性能和更低的电容。一个更典型的情况是要求核心芯片能适应 3.3V 和 1.5V 的 I/O，需要采用多个不同的 I/O 标准。设计者只能从下述两种方案中选用一种：使用可以支持更高电压的氧化层厚度同时能容忍低电压标准的性能问题，或使用较薄的氧化层器件同时采用复杂的结构以避免器件受到 3.3V 标准的过电压条件。氧化层较厚的器件所具有的较高阈值电压和接口电路要求的较低电源电压会引起性能问题。图 6-3 是参考文献［4］中介绍的一种层叠输出缓冲器的简化电路图。当时开发的这种结构是为了用 1.8V 的晶体管支持 1.8V、2.5V 和 3.3V 的 I/O 标准。层叠 I/O 结构可以采用较薄氧化层器件支持多 I/O 标准，具有改善 I/O 性能的潜能。通过给 P_{bias} 与 N_{bias} 提供适当的偏置，可以产生一个受控的阻抗，实现板级信号和负载的匹配，减少反射，改善信号完整性。利用层叠结构的优点并采用一组器件实现期望的阻抗，可以改善输出阻抗控制。这种方式存在的问题包括热插孔保护、ESD 保护和不使用缓冲器时的三态输出。

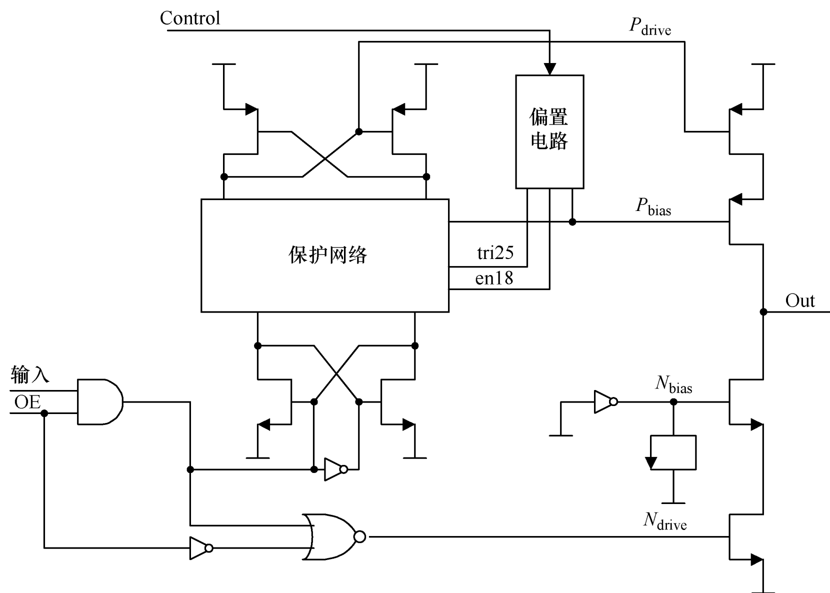


图 6-3 层叠输出缓冲器框图（来源于参考文献［4］）

6.3.2 差分缓冲器

差分缓冲器是真正高速数据转换器的唯一选择。图 6-4 所示为共模反馈 LVDS 输出缓冲器的简图[13]。电阻 R_1 和 R_2 用来将输出缓冲器的共模电平提供给反馈放大器，调整 MN_1 与 MP_1 的偏置电压以保持共模电压为常数。参考电压一

般由带隙基准产生，能实现缩放以提供期望的共模电压。采用带隙基准能减小共模变化以及电源噪声敏感度。这个技术已在实际设计中得到普遍使用，因为反馈环路提供了很好的共模电压控制。R_1 和 R_2 必须很大，以避免给输出缓冲增加负载。如果有可能，给两个电流源（MP_1 与 MN_1）增加级联。由于具有优越的电流匹配，可以进一步改善缓冲器的性能，并且可以增加输出阻抗。对缓冲器的整体性能来说，确保两个电流源（MP_1 与 MN_1）相等非常重要。关于使器件失配最小化的细节请参看第 10 章。两个电流源之间的失配会导致输出信号的共模电压漂移。

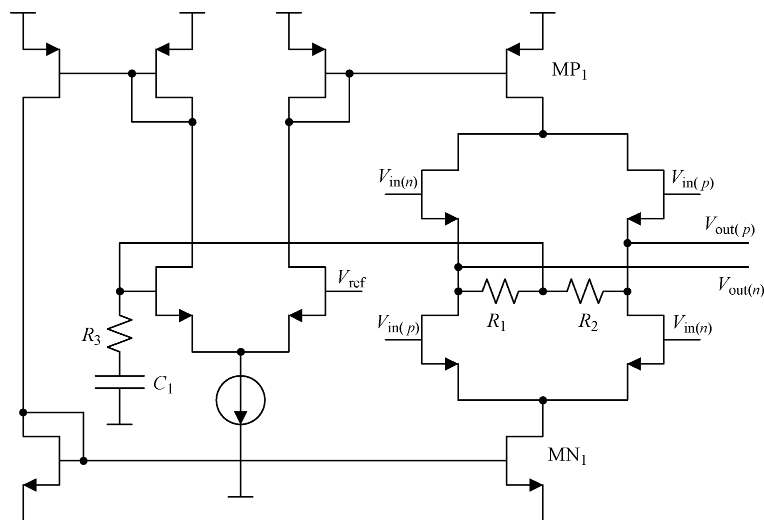


图 6-4　采用有源共模反馈的 LVDS 缓冲器简化电路图
（来源于参考文献［13］）

电流模式逻辑缓冲器　由于 CML 缓冲器表现出的高性能，已在许多应用中获得广泛的认可，包括最高速的串行接口，因为在标准的 CMOS 工艺中它们有高达 10Gbit/s 的速度并且还可能更高[3]。在 90nm 或更小的 65nm 工艺中，这种趋势并没有改变。电流模式逻辑可以采用交流耦合或者直流耦合，而交流耦合一般更常用。图 6-5 所示为一个逐级放大的 CML 输出缓冲器简图。使用逐级放大是因为最后一级一般要驱动较大的负载。为了减小缓冲器的时延，采用了类似于反相器链的串连比例连接模式。当每级的时延相等时，总的时延将最小[11,12]。在位于中间的各级插入电感，通过时延流过电阻的电流使电容的充电速度更快，从而可以增加信号沿变化率[6]。这些电感值的范围为 2～10nH。这些电感有可能要求很大，很难实现，因此经常使用有源负载来起到电感的作用。位于一个器件的栅与相对器件漏之间的耦合电容用来减小栅漏重叠电容效应，这个效应会使输入信号耦合到输出节点。因此添加这些电容有助于减小该效应。给定了这种类型缓

冲器的典型速度要求后，一般会使用 NMOS 器件，因为它们的驱动能力强。这些器件必须一直处于饱和状态以确保高速性能。

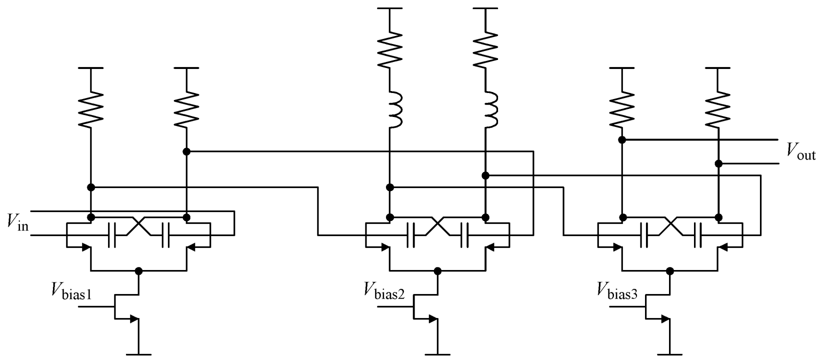


图 6-5　三级 CML 输出缓冲器简图

图 6-6 所示为一个驱动采用交流耦合负载的 CML 缓冲器最后一级的简图。采用交流耦合可以使不兼容的电压电平能连接在一起。如果使用交流耦合，必须使用例如 8B/10B 标准对数据编码，以阻止共模电压漂移。如果接收器对共模电压的时延敏感，共模电压的任何变化将表现为接收信号的抖动。

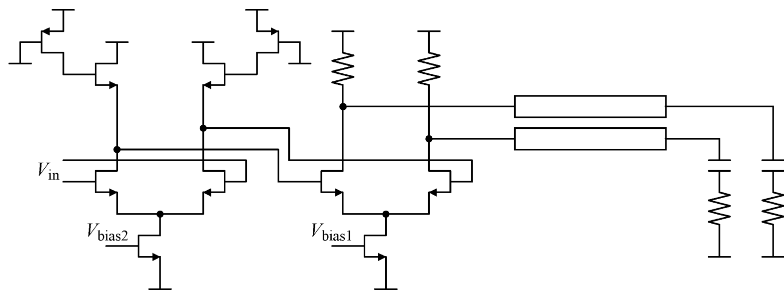


图 6-6　驱动采用交流耦合的片外负载的 CML 缓冲器

图 6-7a 所示为多电平脉冲幅度调制方案简图，通过对每个传输的四个电平进行编码，增加输出和输入缓冲器的有效带宽。这个波形是将多位转换互相叠加形成的。这种方式可以使数据率有效加倍而不必使频率加倍。当然，这个方案使得输入和输出缓冲器变得更加复杂了，因为有四个独立的电平必须被发送和接收，而不是像前面讨论的两电平方案。时钟恢复单元变得更加复杂，因为这四个电平必须被编码以决定输入信号相对于参考时钟的相位。这种信号方案最大的缺点之一是尚没有一个普遍认可的标准，使得该技术的发展很困难。图 6-7b 是一个采用 Alexander 相位检测器的 4 脉冲振幅调制接收器简图，是时钟数据恢复系统的一部分。采用一个两位模数转换器对数据进行采样，以确定有效电平，将其

解码为两位数据。

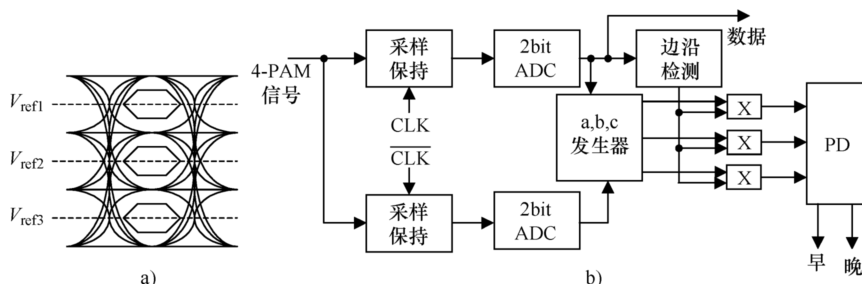


图 6-7　多电平脉冲振幅调制信号波形和系统图

6.4　ESD 保护

ESD 已经在第 5 章做过分析，因此这里仅作简要讨论。对较厚氧化层器件来说，这些先进工艺中的 ESD 保护问题没有变化。用于 0.5μm、0.35μm 和 0.25μm 工艺的保护方案仍然有效。ESD 保护方案的主要变化集中于对采用薄氧化层器件的 I/O 的保护，因为它们对 ESD 更加敏感。第二个目标是如何减小输入和输出缓冲器的附加电容以改善整个系统的速度。确保电容不会随着所加电压的变化而改变也非常重要。否则，时延会随着电压变化，引起信号的额外抖动。

图 6-8 所示为如何设计输出缓冲器以改善 ESD 性能的简图[9]。该结构没有使用最小栅到接触间距，而是利用远大于最小设计规则的间距，多数情况下间距在 1μm 以上。这个技术有效地增加了接触孔到漏和源的电阻，这个电阻起着限流的作用以改善器件的 ESD 性能。图中器件源和漏两侧均有电阻。一些研究表明漏侧的电阻对器件保护来说就足够了，因为串联电阻减小了漏电压，这最终减小了栅上的应力。这个电阻在二次击穿的回压期间起着减小漏极电流的作用[15]。但是，增加接触孔到栅的间距会增加器件的输出电容，从而使输出缓冲器的性能退化。这个方式的第二个缺点是输出缓冲器占据了很大的面积。最后一个缺点是要求有一个无金属硅化的区域，这就要求增加掩膜层数。

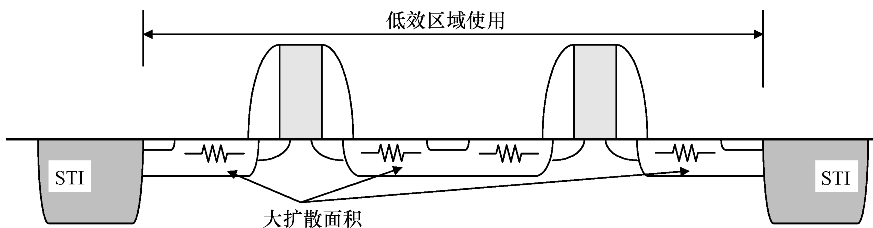


图 6-8　采用双“指条”的输出缓冲器晶体管简图

保护输出缓冲器的第二种方式是使用参考文献［14］中介绍的方法，即在输出压焊点上添加一个晶闸管（SCR）结构，保护连接到压焊点的器件免受 ESD 损伤。只要 SCR 结构的触发电压小于输出器件的击穿电压，这个方案将非常有效。采用这个方案，输出器件就可以设计得非常小，从而减小引脚电容。采用低电容 SCR 结构可以进一步减小引脚电容。

6.5 I/O 开关噪声

在这些先进工艺中，由于更多的 I/O 接口数、增加的边缘变化率和增加的工作频率，I/O 缓冲器的开关噪声变成了一个相当严重的问题。前几代的设计可以依赖电路板级的去耦，但是现在的高速 I/O 要求在晶片上实现去耦以满足现代先进系统的严格要求。这一点如图 6-9 所示。图中显示了一个高速接口的

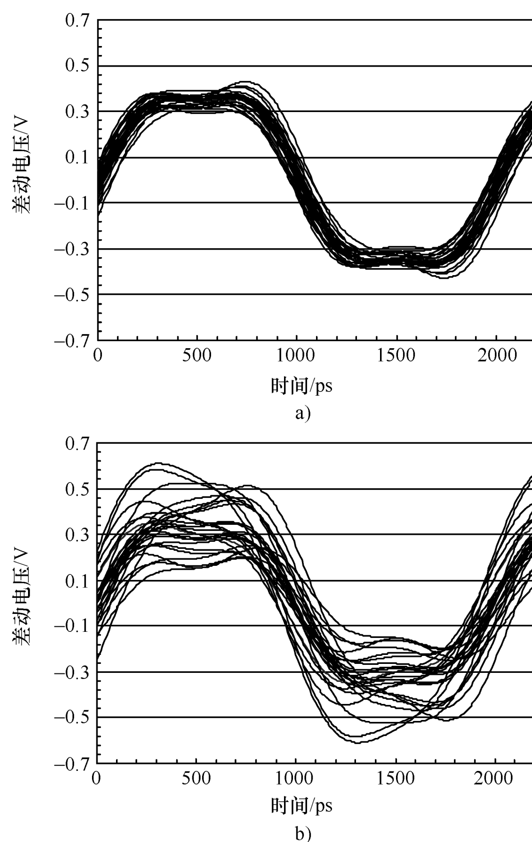


图 6-9　16 位宽、最大速率为 1Gbit/s 的并行接口仿真结果

a）有片上去耦　b）无片上去耦

输出，当其余 15 个 I/O 采用伪随机位组合触发时，该高速接口能够采用最大数据率（1Gbit/s）触发。如果未包括片上去耦，显而易见会出现码间干扰。没有片上去耦，电源电压降低和地弹会变得非常重要，以至于会引起输出缓冲器失效，特别是所有 I/O 受到从高电平到低电平或者从低电平到高电平转换的驱动时。实现有效的板级去耦变得很困难，因为引脚数很多，很难将去耦电容放置到与器件足够靠近的位置，特别是需要放在片外时。另外，即使在器件封装上作出了重大的改进以减小寄生电阻和电感，对当前的 I/O 需求，这仍然是不充分的。总之，必须不断强调的是，在这些先进工艺中，如果要求 I/O 必须支持各种高速接口，那么芯片上的 I/O 去耦就是绝对必须的。

图 6-10 所示为一排有 8 个 I/O 的 I/O 组合与相关寄生元件的模型简图，其中忽略了封装寄生电容。此图经过扩展，可以对采用倒装焊封装的有 68 个 I/O 的实际 I/O 组合进行仿真。I/O 开关对 I/O 处电源电压的影响如图 6-11 和图 6-12所示。这个模型将 I/O 分成 4 簇，以简化仿真。封装和芯片寄生参数是基于 90nm 工艺估计的。每个 I/O 采用不同的去耦量以描述片上去耦的重要性。两个图显示的是每个 I/O 增加 9pF 的结果。图 6-11 所示的平均电源电压曲线说明了电源总线未能精确平衡，这也解释了为什么尽管电源差别很小，即平均电源电压几乎为常数，但是一个终端的 *IR* 压降大于另一个。图 6-12 所示的最大电源电压变化是通过在每一个缓冲器组（四个缓冲器一个组形成一个簇）计算 V_{CC} 和 V_{SS} 电平差值，并确定最大偏差。表 6-1 总结了不同去耦程度对 I/O 组电源电压的影响。

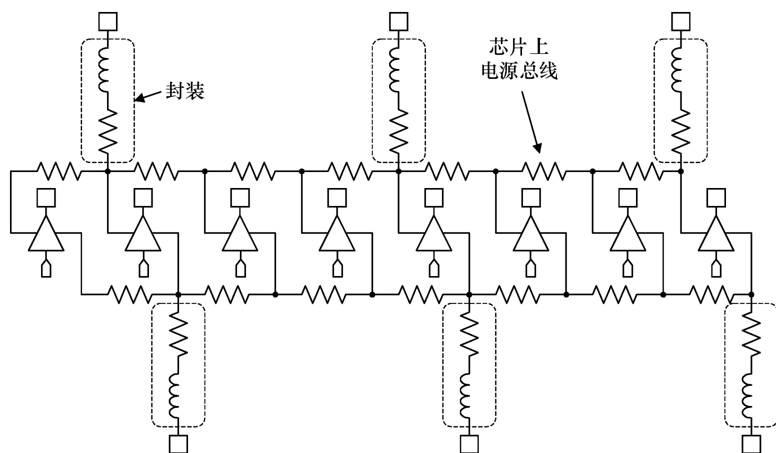


图 6-10　包括了寄生元件的一排 8 个 I/O 的简图

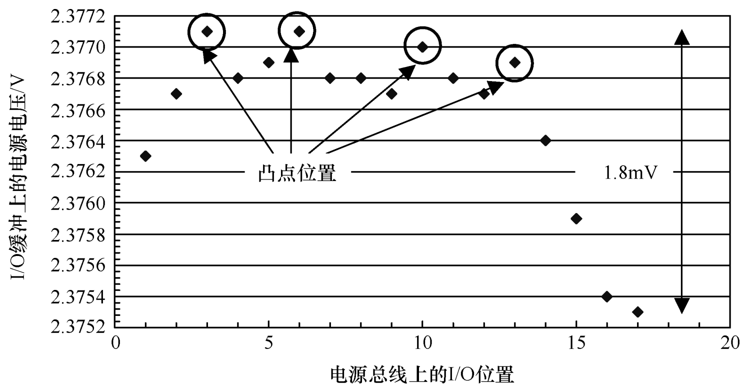


图 6-11　采用 100MHz 频率开关时 I/O 上的平均电源电压

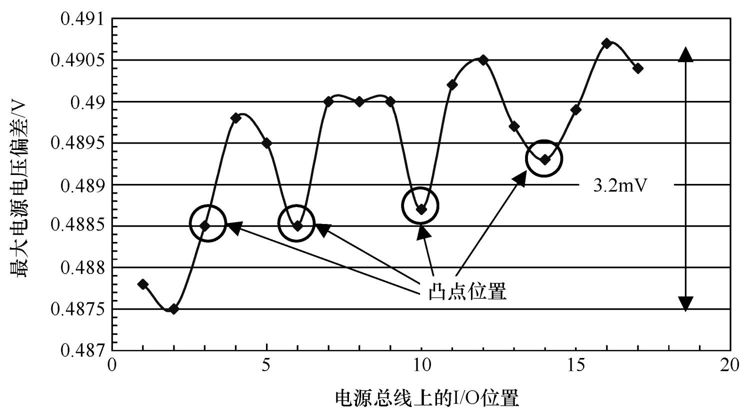


图 6-12　半数 I/O 发生开关作用时的最大电源电压变化（具有 68 个 I/O 的组中，每 I/O 缓冲器采用 9pF 去耦）

表 6-1　电源电压变化与去耦程度关系的总结

每个 I/O 的去耦/pF	平均电源电压/V	最大电源电压偏差/V
4.5	2.38	1.03
9.0	2.38	0.49
18.0	2.38	0.26

基于这些仿真明显可见，一个 I/O 组上局部的感性和阻性压降会变得很重要。即使采用低感性封装，片上去耦仍是使 IR 压降最小化的唯一有效方法。为了成功设计一个 I/O 系统，推荐采用下列步骤。

（1）评估每个 I/O 的抽取电流。获得抽取电流的正确时序以得到精确的建模是非常重要的。应该对每个工艺角都做这些工作。

（2）决定 I/O 主电源线可用的金属层的多少（这一般是基于上一代工艺电源线获得的），为 I/O 主电源线形成一个实际的电阻网络。这一步必须与第一步结合使用以确保不与电迁移规则相冲突。在这个阶段应该监视电迁移以确保各个级别（晶片上、凸点、封装和焊球）都有合适的主电源线。

（3）评估在凸点和 I/O 电源网格间形成阻性连接的 I/O 电源总线凸点走线。这时假定采用倒装芯片封装，但是该方法对金属线键合封装同样适用。

（4）开发/获得凸点模型。凸点的电阻范围为 0.01 ~ 0.1Ω，电感约为 0.1nH，因此它们不能被忽略。

（5）评估封装电阻网络。即使在封装中使用等位面图形（plane），仍然有一定的电阻和电感。

（6）开发/获得封装焊球的模型。焊球的电阻范围为 0.01 ~ 0.1Ω，电感约为 0.2nH 以上，因此它们不能被忽略。

一旦完成这些任务，就必须进行折衷研究以观察增加或减小 I/O 电源总线、增加凸点、减小凸点与电源总线之间的走线、增加封装走线/plane 和添加附加焊球等的影响。基于这一分析可以确定凸点和焊球之间的相对位置。另外，可以在增加附加的片上去耦电容、电源总线和封装凸点和焊球之间进行折衷。为了得到适当的折衷，这个分析必须在项目启动之初进行。

参考文献［16］介绍了考虑电源噪声增加引起负反馈的情况下，I/O 开关噪声效应的一般模型。其中电源噪声可以由下式得到：

$$V_n = V_k + \frac{T}{L_p}\frac{p}{nK}\left(1 - \sqrt{1 + 2V_k\frac{L_p nK}{pT}}\right)$$

式中，$V_k = V_{CC} - V_{in}$；T 为输出缓冲器达到最大值时的上升时间；L_p 为每个引脚或焊球的电感；p 为引脚或者焊球数；n 为同时开关的输出缓冲器数；$K = \mu_n C_{ox}(W/L)$。这个简单的方程并不能给详细设计 I/O 电源组提供必要的精确度，但是这个方程的一般形式对确定多个 I/O 同时开关的电源噪声效应非常有用。对方程做修正，可以得到：

$$V_n = A + \frac{B}{n}(1 - \sqrt{1 + Cn})$$

然后使用 SPICE 仿真确定系数 A、B 和 C，可能得到一个确定 I/O 开关噪声效应的简单表达式。

6.6　匹配

对于单个片上的巨大系统，匹配已经变成一个主要的问题。图 6-13 所示为一个单端端接方案简图。该端接被用来实现驱动器、传输线和负载之间的阻抗匹

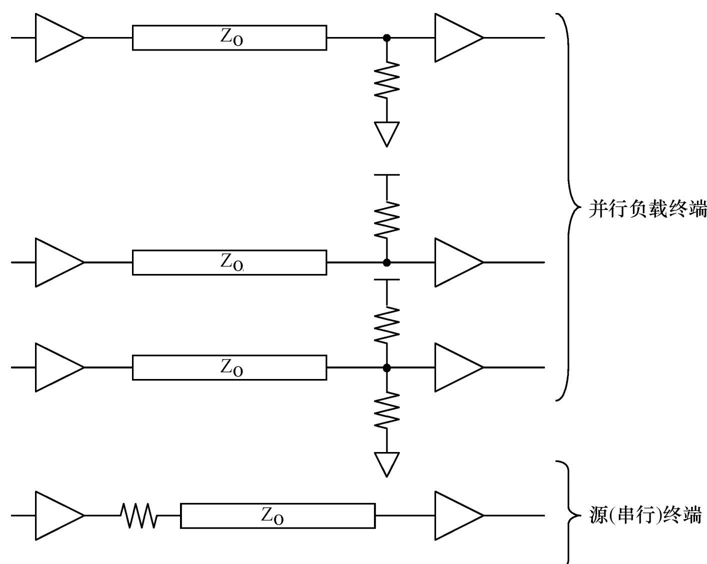


图 6-13 单端端接方案

配，使得由阻抗不连续引起的反射最小化。这些反射将引起码间干扰。对一定规模的 ASIC，有几百个 I/O 引脚，在电路板上给这些高速接口提供合适的匹配，即使不是不可能的，也极其困难。因为这些引脚密集的元件附近非常拥塞，高 I/O速度和剩余的封装与板级残余会引起另外的中断。这些限制迫使 IC 设计者在晶片上包含有匹配，特别是对速度非常高的接口。这种趋势在每个工艺节点都变得越来越强烈。图 6-14a 所示为一个差分输入缓冲器可能的输入端接方案简图。匹配电阻被放置在差分输入缓冲器的输入之间。数字总线被用来设定并联电阻的值以获得期望的端接电阻值。因为它们对电压的低敏感性、受控的寄生电容和适当的精度，用未硅化的多晶形成电阻，虽然它们的变化范围仍然在 ±20%。因此，采用并联的电阻方式来补偿工艺和温度的变化。这种实现方法意味着必须采用几种电路来检查并补偿这些因素。一种经常采用的设计是利用外部电阻得到参考电压，包括一个如图 6-14b 所示的片上端接电路形式。虽然这个方法有效，但是它要求两个额外的引脚并且将外部元件的精度耦合到整体的电路性能中。第二种方法是使用微调以产生精确的基准。这种方法可以不增加额外的引脚，但是增加了相当的复杂度，并加大成本，因为微调方案必须使用几种熔丝技术（金属或多晶）。另外，当增加片上端接时，必须监测下述关键参数：

（1）工艺、电源和温度变化。这些参数的敏感度必须至少最小化到 ±10%。

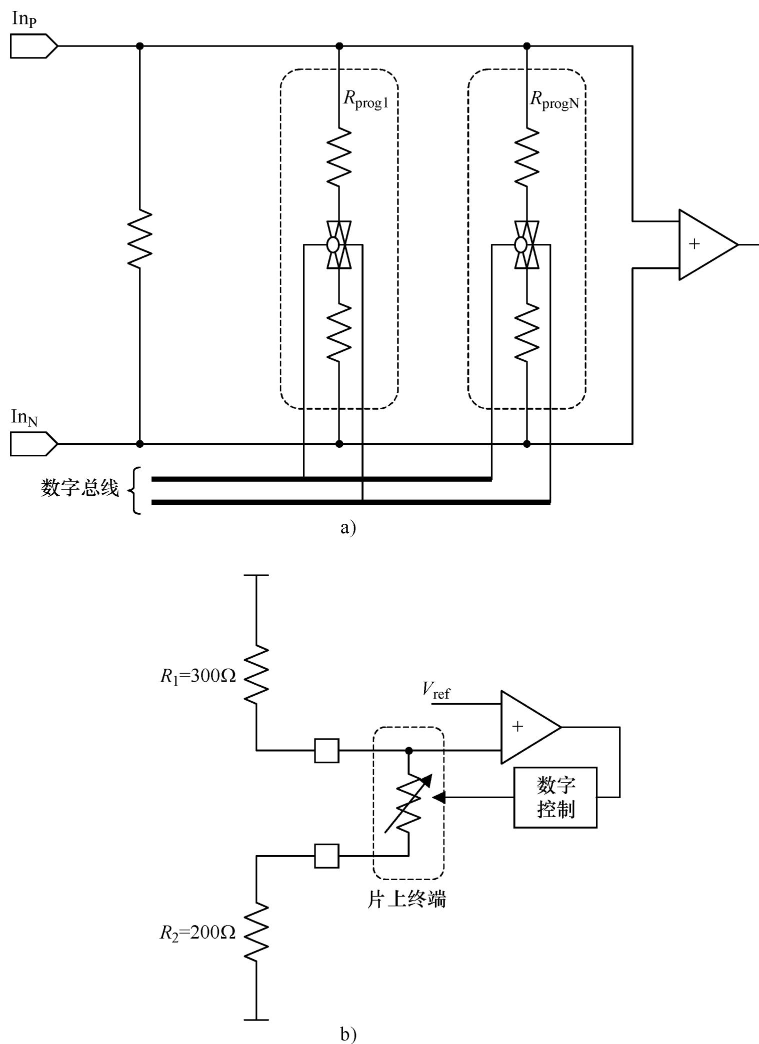


图 6-14　a）可变输入端接方案简图　b）校准电阻设定所要求的电路简图

（2）附加的引脚电容。必须使该参数最小化；否则，它将限制接口的最大工作速度。

（3）电容与电压的函数关系。必须使电容与电压的关系最小化，并且变化量应该小于总电容的 5%，以确保电压对电容的影响不会严重影响接口的性能。

提供一个精确的端接阻抗以维持与驱动源和传输线的匹配是非常重要的。使端接阻抗具有可配置的特点也正变得更加重要，这样终端用户可以根据自己的特殊用途选择合适的端接。

6.7 阻抗匹配

当输出缓冲器驱动传输线时，为了保证良好的信号完整性，实现阻抗匹配非常重要。如果不能实现阻抗匹配，发生的反射将使整个系统的性能退化。这个问题仅对先进的工艺非常重要，因为其要求有较高的数据速率，而 I/O 时序开销却不断减少。已经有几个方案可以提供可编程输出阻抗[7,8]。图 6-15 所示为具有这些功能的一个电路实例。六个控制信号 ϕ_1 到 ϕ_6 被用来设置缓冲器的输出阻抗。当单个器件开启时，同样的方案也可以与逐级放大一样被用来控制压摆率。

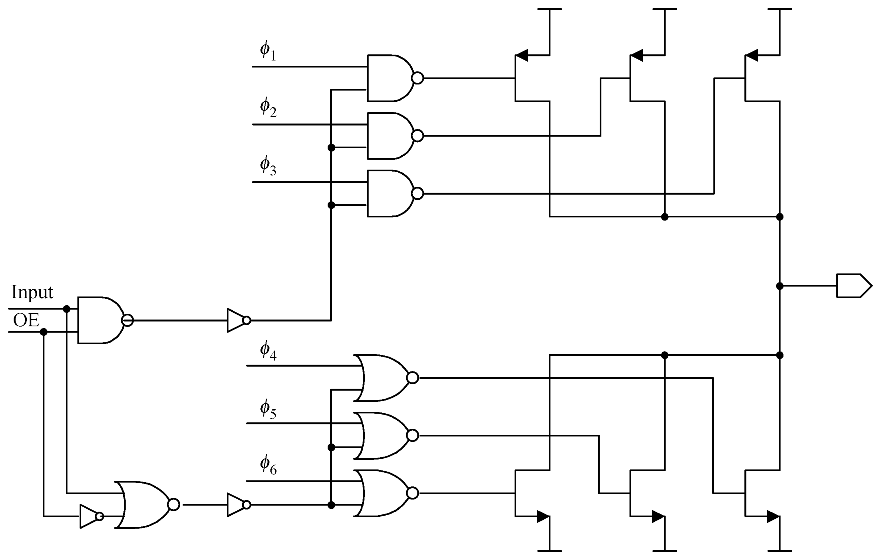


图 6-15 输出缓冲器阻抗控制电路简图（来源于参考文献［7］）

这个电路不是没有问题，因为驱动器的输出阻抗会随着工艺、温度和电源电压而剧烈变化。解决这种变化影响的唯一方法是采用几种校验电路抵消这些效应。应该评估工艺、温度和电源对整个信号完整性的影响，以确定为了满足系统时序规范要求，是否需要采用校准电路。

6.8 前置放大

有两项技术可以被用来增加 I/O 的工作频率：前置放大和均衡化。对高于 1Gbit/s的传输速率，采用这两项技术正显得越来越重要，而对 1.5Gbit/s 以及更高的传输速率，这两项技术是绝对必须采用的。近几年，主要由于更高性能的核心逻辑要

求 I/O 具有更宽的带宽，使得前置放大已经从学术研究转向为广泛用于主流产品生产中。前置放大是这样一个概念：在发生信号跃变的时刻即提供一个过驱动电流。

图 6-16 和图 6-17 所示为前置放大的概念的说明。图 6-16 是一个发射机、接收机和可以由 PCB、连接器或电缆实现收发之间互连的简图。图中发射芯片没有采用前置放大，发射的波形如发射机电路下方所示。这种情况下接受端的最终波形如接收机芯片下方所示。由于更高频率下出现更大的损耗，接受的波形显示了 PCB 或电缆损耗的影响，因为它们扮演着低通滤波器的角色。

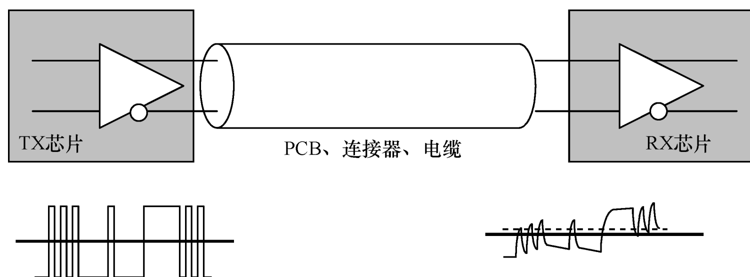


图 6-16　无前置放大的两个芯片之间的高速接口

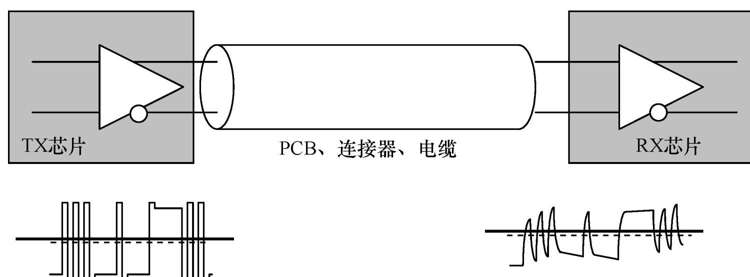


图 6-17　有前置放大的两个芯片之间的高速接口

图 6-17 中的系统配置与图 6-16 所示的相同，但是其中发射芯片采用了前置放大。在数据从低到高或者从高到低跃变的每一点上都产生附加的驱动电流。这个增加的驱动电流有助于增加接收机的电压摆幅，使接收机更容易恢复数据。

可以采用类似于图 6-18 所示电路实现前置放大控制。当跃变发生时，在串行器中对数据进行采样以确定什么时刻出现数据跃变，并且产生一个控制信号，使输出缓冲器受到额外电流驱动，如图 6-19 所示。仅当数据发生跃变时，MP_2 和 MN_2 两个晶体管被使能以有效增加缓冲器的驱动电流，这可以增加电压摆幅。这个简单实例显示的是使用单个器件控制前置放大，但是实际上，会采用多个器件用于对前置放大程度进行配置，以满足器件的特定应用需要。当设计前置放大方案时，必须考虑工艺、电源和温度的影响，以保证电路可以提供合适的跟踪，维持良好的信号完整性。

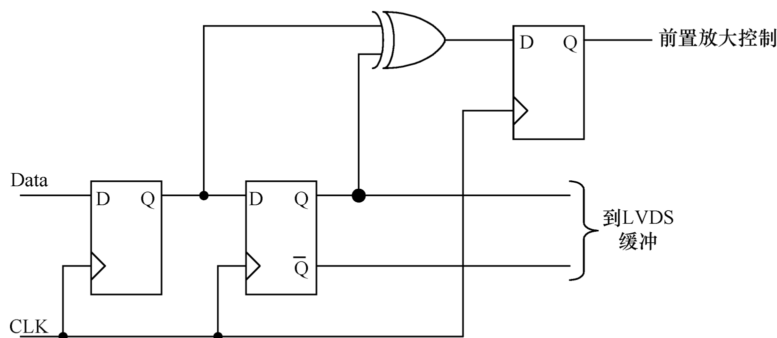


图 6-18　产生前置放大控制信号的电路简图

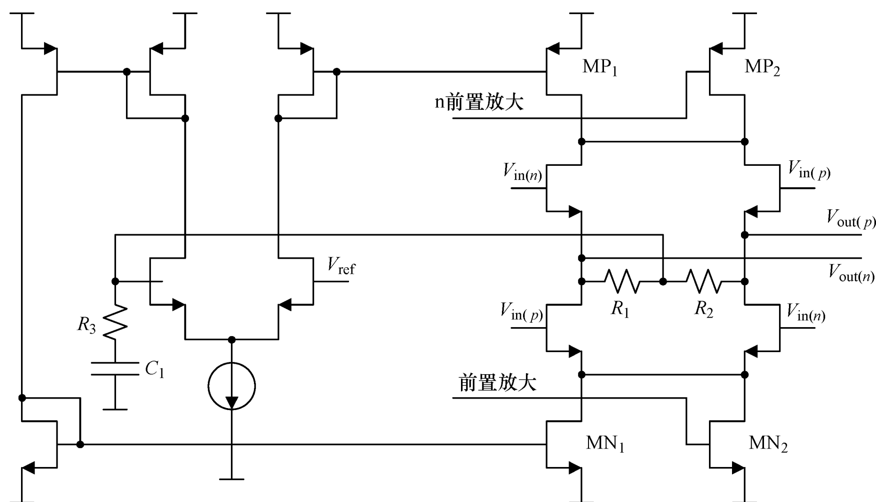


图 6-19　增加了前置放大的 LVDS 输出缓冲器简图

6.9　均衡化

前置放大技术用在数据链的发射一侧，而均衡技术则被用在数据链的接收一侧。均衡化业已成为任何高速输入缓冲器设计中不可缺少的一部分。同时使用均衡化和前置放大为满足高速时序要求提供了最好的整体解决方案。图6-20所示为实现均衡化的电路简图。均衡化方案是通过采样保持电路监测当前数据（$V_{in(n)}$）和前一个数据（$V_{in(n-1)}$）。图 6-20 中的 MN_1 的漏电流可以由下式给出

$$I_{DMN1} = g_{mMP1}[V_{in(n)} - V_{ref}] - g_{mMP3}[V_{in(n-1)} - V_{ref}]$$

记 $\alpha = g_{mMP3}/g_{mMP1}$，可将上式简化为

$$I_{DMN1} = g_{mMP1}\{[V_{in(n)} - V_{ref}] - \alpha[V_{in(n-1)} - V_{ref}]\}$$

图 6-20　均衡化器简图（来源于参考文献［1］）

通过缩放尾电流 I_1 和 I_2 或者调整 MP_3 和 MP_4 相对于 MP_1 和 MP_2 的尺寸可以改变参数 α。最好使参数 α 为可变化的，从而可以针对特殊的应用，优化确定 α 数值。研究表明，均衡化的引入可以使数据速率增加约 10%[2]。

6.10　总结

随着整个系统对数据率的要求不断增加，基于更先进工艺的 I/O 设计变得更加困难。设计者必须对最终应用有一个完整的理解，以确保能满足严格的要求。必须采用如前置放大和均衡化等新的电路技术来满足更高的 I/O 带宽要求。可以在 CML 缓冲器中使用感性峰值以进一步增加带宽。所有的这些技术增加了 I/O 缓冲器设计的复杂性。

成功实现更高的数据率会要求有更快的边缘速率，而这会同时产生更多的瞬态开关噪声。现在设计者只好采用晶片上去耦，以有助于减小电源电压降低和地弹的影响。必须认真考虑集成在封装外壳内的晶片，以使封装寄生最小。不要单独考虑 I/O 缓冲器的设计，而应该考虑整个 I/O 系统，以使得在维持较小晶片总面积的情况下，使整个 I/O 系统得到完全的优化，提供期望的性能。在启动详细设计之前，应该建立初始的宏模型以确立整个系统从晶体管级到系统级的全局要求。

I/O 设计的未来趋势没有发生任何变化，仅数据率在增加。由于串联接口能在消耗更小的总体功耗和采用较少 I/O 引脚数的情况下实现高速工作，正得到更加普遍的使用。这些串联接口的更高的数据率将要求更高的集成度，以保证能实

现性能目标。另外，由于 IC 最终用户要进行系统级仿真以验证他们的板级和系统级设计，因此建模将变得势在必行。

参 考 文 献

[1] J. Y. Sim, J. J. Nam, Y. S. Sohn, H. J. Park, C. H. Kim, and S. I. Cho, A CMOS transceiver for DRAM bus system with a demultiplexed equalization scheme, *IEEE J. Solid-State Circuits*, Vol. 37, pp. 245–250, Feb. 2002.

[2] B. S. Song and D. C. Soo, NRZ timing recovery technique for band-limited channels, *IEEE J. Solid-State Circuits*, Vol. 32, pp. 514–520, Apr. 1997.

[3] M. W. Allam and M. I. Elmasry, Dynamic current mode logic (DyCML): a new low-power high-performance logic style, *IEEE J. Solid-State Circuits*, Vol. 36, pp. 550–558, Mar. 2001.

[4] H. Sanchez, J. Siegel, C. Nicoletta, J. P. Nissen, and J. Alvarez, A versatile 3.3/2.5/1.8-V CMOS I/O driver built in a 0.2-μm, 3.5 nm Tox, 1.8 V CMOS technology, *IEEE J. Solid-State Circuits*, Vol. 34, pp. 1501–1511, Nov. 1999.

[5] R. Mooney, C. Dike, and S. Borkar, A 900 Mb/s bidirectional signaling scheme, *IEEE J. Solid-State Circuits*, Vol. 30, pp. 1538–1543, Dec 1995.

[6] H.-M. Rein and M. Moller, Design considerations for very-high-speed Si bipolar IC's operating up to 50 Gb/s, *IEEE J. Solid-State Circuits*, Vol. 31, pp. 1076–1090, Aug. 1996.

[7] T. Matano, Y. Takai, T. Takahashi, Y. Sakito, I. Fujii, Y. Takaishi, H. Fujisawa, S. Kubouchi, S. Narui, K. Arai, M. Morino, M. Nakamura, S. Miyatake, T. Sekiguchi, and K. Koyama, A 1-Gb/s/pin 512-Mb DDRII SDRAM using a digital DLL and a slew-rate-controlled output buffer, *IEEE J. Solid-State Circuits*, Vol. 38, pp. 762–768, May 2003.

[8] T. Takahashi, M. Uchida, T. Takahashi, R. Yoshino, M. Yamamoto, and N. Kitamura, A CMOS gate array with 600 Mb/s simultaneous bidirectional I/O circuits, *IEEE J. Solid-State Circuits*, Vol. 30, pp. 1544–1546, Dec. 1995.

[9] M. D. Ker and T. S. Wu, Novel octagonal device structure for output transistors in deep-submicron low-voltage CMOS technology, *IEEE International Electron Devices Meeting*, pp. 889–892, 1996.

[10] F. A. Musa and A. C. Carusone, Clock recovery in high-speed multilevel serial links, *IEEE International Symposium on Circuits and Systems*, Vol. 5, pp. 449–452, May 2003.

[11] P. Heydari, Design and analysis of low-voltage current-mode logic buffers, *IEEE International Symposium on Quality Electronic Design*, pp. 293–298, Mar. 2003.

[12] P. Heydari and R. Mohavavelu, Design of ultra high-speed CMOS CML buffers and latches, *IEEE International Symposium on Circuits and Systems*, Vol. 5, pp. 208–211, May 2003.

[13] A. Boni, A. Pierazzi, and D. Vecchi, LVDS I/O interface for Gb/s-per-pin operation in 0.35-μm CMOS, *IEEE J. Solid-State Circuits*, Vol. 36, pp. 706–711, Apr. 2001.

[14] M. D. Ker and C. H. Chuang, Electrostatic discharge protection design for mixed-voltage CMOS I/O buffers, *IEEE J. Solid-State Circuits*, Vol. 37, pp. 1046–1055, Aug. 2002.

[15] T. L. Polgreen and A. Chatterjee, Improving the ESD failure threshold of silicided NMOS output transistor by ensuring uniform current flow, *IEEE Trans. Electron Devices*, Vol. 39, No. 2, p. 379, 1992.

[16] R. Senthinathan and J. L. Prince, Simultaneous switching ground noise calculation for packaged CMOS devices, *IEEE J. Solid-State Circuits*, Vol. 26, No. 11, pp. 1724–1728, Nov. 1991.

第 7 章 DRAM

7.1 引言

在过去的几十年里，制造厂商已经成功的优化并缩小了动态存储器（DRAM）。成功实现 DRAM 缩小的关键技术是三维单晶体管/单电容单元和锁存交叉耦合读出放大器。然而，进入 100nm 工艺时，DRAM 采用的这些专用电路正面临着特殊的挑战，使得人们业已熟悉的那些原理将不得不被摒弃。本章将讨论这些特殊的挑战。为了便于分析，首先简要介绍有关 DRAM 的一些最重要的基本内容。

7.2 DRAM 基础

在动态存储器中，数据以易失电荷的方式存储在电容中。由于易失的特性，如果每个存储器单元不能得到定期刷新，存储电荷将会流失。在早期 3T1C 单元[1]的基础上，出现了优化的 1T1C 单元[2]。1T1C 是英文全称的首字母缩写，表示存储单元中包括一个用于存取信息的晶体管和一个用于存储信息的电容。

图 7-1 所示为一个存储器阵列中存储器单元分布的电路图。字线与排成一列的存取晶体管的栅极相连；位线与字线垂直，与排成一行的源极相连。普通电容的理想化版图如图 7-2 所示，其中没有包括关于如何在硅上实现这种结构的细节。图中反映的相关信息只是每个这种单元版图需要 $8f^2$ 的面积，其中 f 是工艺的最小特征尺寸。面积为 $8f^2$ 的单元在给定排列下可以高效地实现折叠式位线结构[3]。这种结构中，位线被安排为互补对，位线（t）与互补位线（c）交替排列。字线连接到位于位线或互补位线上的单元（见图 7-1）。

现代绝大多数 DRAM 中（即使不是全部）选择的是折叠式位线结构。在折叠式位线结构中，在位线的每个互补对中有一个精确的差分放大器——读出放大器。当以最近的距离放置差分对时，大多数的阵列噪声以共模噪声的形式出现，从而不会导致差分放大信号的退化。

图 7-3 是一个典型的读出放大器电路图。为了实现面积最小化，一个读出放大器被其左右两个存储阵列共享。通过多路复用器选择被读出的阵列。均衡器晶体管使读出放大器偏置到所期望的电压值为 VAA/2 的期望的预充电电平 ，其中

VAA 是位线电压的高电平。最后，位于中间的交叉耦合对是读出放大器的核心。

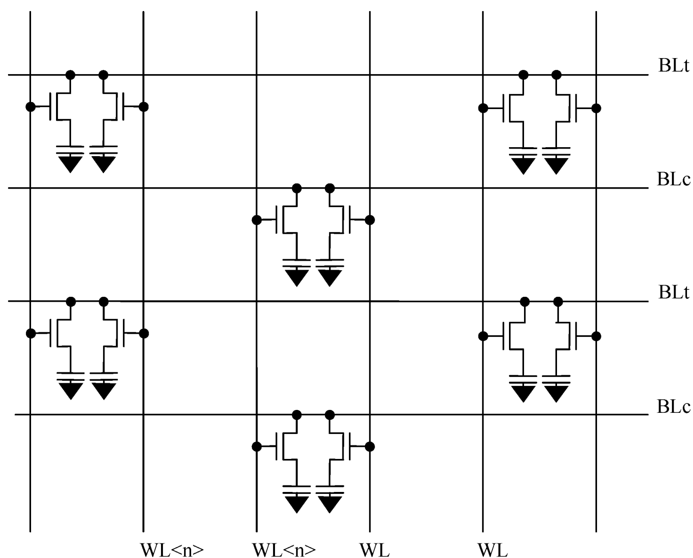


图 7-1　DRAM 阵列电路示意图：垂直方向的是字线（WL），水平方向的是位线（BL）。逻辑上，位线由位线/互补位线对组成

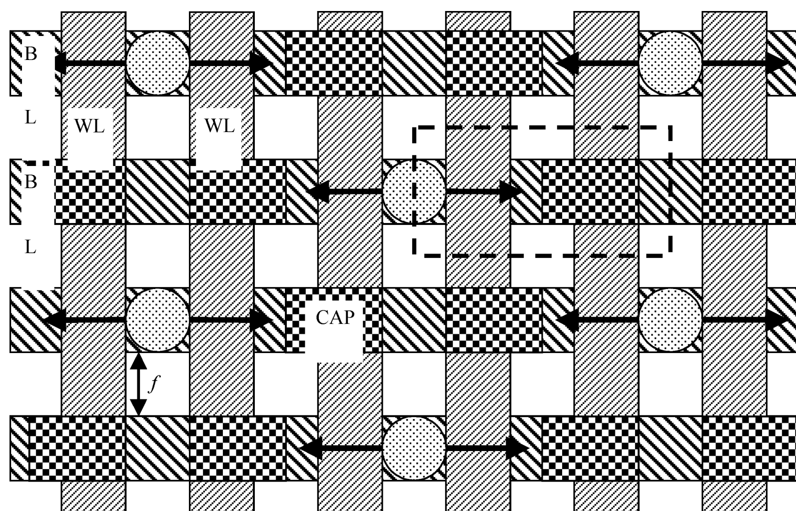


图 7-2　面积为 $8f^2$ 的 DRAM 阵列的理想版图：虚线框为存储器单元的轮廓；箭号代表传输器件中的电流方向；圆表示位线与传送器件扩散区的接触；字线（BL）为垂直走向；位线方向与其垂直。存储电容（CAP 方格盘型图案）面积可能超过 $1f^2$

图 7-3　经典的共享 DRAM 读出放大器的电路图

（为清晰起见，图中仅显示了一条数据存取器件路径，其没有缩放限制。多路器和预充电器件在左、右阵列之间选择。这样，一个读出放大器可以用于两个相邻的存储器阵列。Nt 和 Nc 是读出放大器 n 型晶体管，用于实现位线或互补分支的下拉。）

图 7-4 所示为一个典型的存储器读取周期。最初，阵列保持在预充电状态。所有的多路器门和均衡器门被偏置到高电平并且位线保持在 VAA/2。作为激活时序的第一步，首先要选择读出的是右侧还是左侧阵列。在本例中，读出的是右

图 7-4　阵列中典型的信号波形图

（图中所示的是读 1 的情况。实线代表位线和互补位线。虚线代表字线信号，点线和点划线分别代表均衡器和多路器信号。MUXL 和 EQLR 相互重叠）

侧阵列。这是通过关闭左侧的多路复用器件（MUXL）和右侧的均衡器（EQLR）达到的。之后，字线会被抬高，存储器单元中存储的电荷转移到位线上。具有均衡电压的位线和存储单元的高（低）电势之间的电荷共享将引起位线电压的增加（降低）。读出后，读出放大器会收到使能信号，放大互补位线对上的差分信号。放大器的使能是通过将 SAN 从 VAA/2 驱动到 0 而将 SAP 从 VAA/2 驱动到 VAA 实现的。这样，存储的信号会自动而正确地被写回到存储单元。这个循环结束后，信号回到它们的预充电电平。

这种读出方案有几个优点。对提供均衡、左-右选择、和放大的电路来说，器件数最少。另外，互连方案也相对简单，可以获得一个较紧致的版图。这对 DRAM 的芯片尺寸来说极其重要，因为 DRAM 中需要的读出放大器非常多。除了具有面积优势之外，从功耗的观点来看，这个电路也是非常有益的。在放大期间，只有即将向上走的位线需要从 VAA/2 充电到 VAA，这个过程消耗的电荷为 CBL × VAA/2（CBL 为位线电容）。在这个周期结束时，电平为 VAA 的高电平位线和电平为 0 的低电平位线通过均衡器晶体管被简单的短路，达到常规的 VAA/2 预充电电平。因此，实现快速均衡，并且不需要附加电流。

这种采用交叉耦合读出放大器的 1T1C 单元实现了高度的面积优化。因此，当工艺缩小进入 sub-100nm 工艺时，仍希望保持这种基本结构。然而，要实现它需要解决多个令人困扰的问题。实际上任何情况都有与自身相关的特殊问题。这些将在下面几节讨论。

7.3　电容的缩放

DRAM 中存储电容是第一个关键元件。通过与位线电容 C_{BL} 的简单电荷共享，加给读出放大器的可用信号 V_{signal} 与存储电容 C_{cell} 之间关系为

$$V_{signal} = \frac{C_{cell}}{C_{cell} + C_{BL}}\left(C_{cell} - \frac{V_{AA}}{2}\right)$$

一般位线电容是存储器单元电容的 5 ~ 10 倍。位线电容包括总的存储器电容加读出放大器的电容。DRAM 单元的设计者努力将各代工艺的存储器单元电容维持在相同的水平。这样，对存储电荷泄露损失的敏感度就有了可比性。电容分配比也保持为常数。如果位线电容随着工艺的发展而降低的话，甚至能够使电容分配比得到改善。

特征尺寸下降到 100nm 工艺时，实现巨大电容的一个关键技术是三维电容。扩展到第三维的可以采用硅上（层叠电容）技术或深入到硅中（沟槽电容）的技术。这两项技术的实例如图 7-5 和图 7-6 所示。

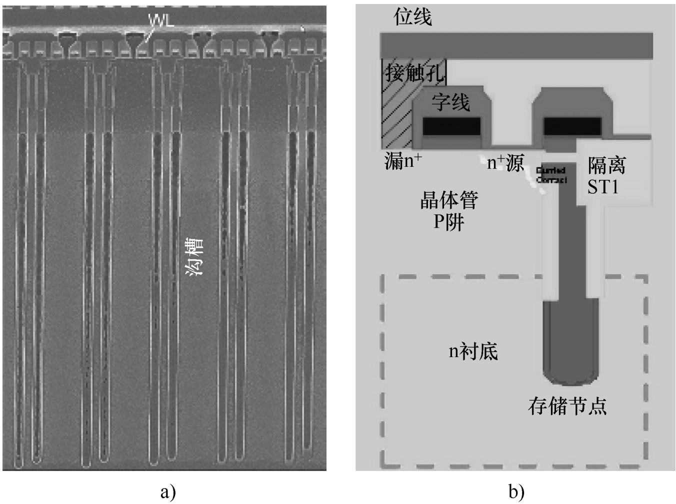


图 7-5 a）沟槽电容单元实例：沟槽刻蚀深入到硅中。注意：与定义特征尺寸 f 的字线的长度相比，沟槽深度比其大数倍 b）字线与相关存储器单元简图

图 7-5a 所示为一个沟槽电容存储器单元，而图 7-5b 所示为沟槽单元的示意图。沟槽刻蚀深入到硅中，这极大地增加了基座上的可用电容面积。图 7-6 是一个层叠电容单元。因为层叠不能做的太高，因此层叠电容结构的面积增强能力小于沟槽方式。如果使层叠太高，对上面介质层进行平坦化处理以便在层叠电容的顶部加工附加金属层将会变得很困难。为了补偿这个缺点，可以采用更进一步的电容增强技术。如图 7-6 所示，电容可以采用粗糙的表面以增加面积。用更高 ε_r 的材料来替代常规的介质层可以进一步提高电容。目前广泛使用的高 ε_r 材料是 Ta_2O_5。

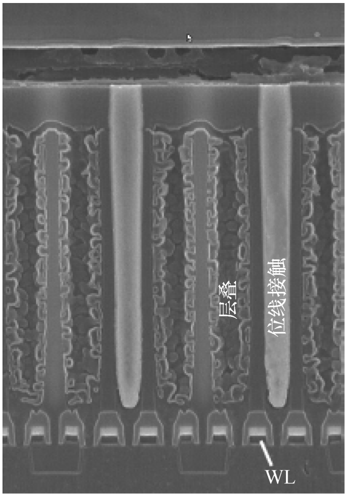


图 7-6 层叠电容单元实例
（注意层叠电容波纹状的表面采用了附加的表面增强技术）

通过持续地改善，这种方法在最近的几代工艺中已经获得了成功。然而，伴随着缩放带来的另一个问题是如何通过电容和（或）接触的串联电阻给单元电容充电和放电。在特征尺寸较小的情况下，刻蚀一个较深的沟槽（见图 7-5）意味着总阻抗的急剧增加。对层叠方式，接触孔从位线向下到硅衬底的长位线存在相似的问题。在阵列的字线之间都需要实现位线和层叠

接触。

这样，直到 100nm 时代，主要重点还是针对如何实现足够大的电容。然而，在未来，如何保证实现足够小的阻抗以保持 RC 常数足够低也将成为必须给予同样关注的重点问题。这将不得不采用低欧姆接触材料和较浅的栅层叠。

另外一种解决问题的努力途径是，保持电容继续缩小，但是使总接触电阻维持在一个电容能被读出的水平。然而，如果单元设计者能够保持电容大致为常数，他们就需要给存储器晶体管提供足够的电流，从而具有足够快速读取单元的能力。因此时下一节的主题就是讨论晶体管的缩放问题。

7.4　阵列晶体管的缩放

我们主要的假定是维持图 7-2 所示的 $8f^2$ 存储器单元。其中，阵列器件的宽度和长度都等于特征尺寸 f。因此，不像逻辑电路中的器件，DRAM 器件的设计中，等比例缩小并不是一个自由的参数。DRAM 器件的缩小必须通过其他方法实现。为了充分理解这一概念，有必要分析阵列器件的边界条件。为了进行更全面的讨论，采用下述符号：VPP 是选中单元的字线电压，VLL 是未选中单元的字线电压，而 VBB 是阵列器件（不再使用 p 型阵列器件）的 p 阱电压。采用三阱工艺，其中阵列 p 阱位于 n 阱中；这可以对传送晶体管施加衬底偏置，从而提升阈值电压 V_{th}，并且也可以将阵列与衬底噪声隔离，并且还能在一定程度上隔离电离粒子的影响。

DRAM 泄漏要求是对阵列器件最严格的限制。DRAM 的标准刷新周期为 64ms。使用 20fF 的存储器电容和可承受 300mV 的泄漏损耗，则最终的总泄漏预算是 100fA。如果将总预算的 10% 分配给器件，器件的泄漏就必须限制在 10fA 之下。由于泄漏电流随着温度升高而增加，因此要求在最高工作温度下器件的泄漏电流也不能超过 10fA。和位线相连的存储单元出现最坏的关态条件是：相应的读出放大器激活，将位线拉低。可以得到

$$I_{leakage}(V_d = \text{VAA}; V_g = \text{VLL}; V_s = 0; V_{bulk} = \text{VBB}; \text{高温时}) = 10\text{fA}$$

然而，器件不能仅仅只针对泄漏进行优化。最重要的是，读出电流必须足够高。简单的说，要求加在单元电容上的初始读出电压斜率为 750mV/2n。在位线被充电到 VAA/2 时的读出条件下必须达到这个值，使器件进入源极跟随模式。由于阈值电压随温度的下降而增加，一般这个条件在低温时是最严格的。器件迁移率随着温度的增加而衰退不能补偿这种阈值电压效应。因此：

$$I_{read}(V_d = \text{VAA}; V_g = \text{VPP}; V_s = \text{VAA}/2; V_{bulk} = \text{VBB}; \text{低温时}) = 10\mu\text{A}$$

我们先总结满足泄漏要求的方法。图 7-7 说明了这种改善。在几代工艺中，许多标准 DRAM 使用负 VBB。采用负 VBB 可以提升器件的阈值电压，因此将减小关态电流泄漏。然而，必须注意 VBB 也是存储节点 n 扩散的 p 阱电压。增加 VBB

将会增大二极管的结泄漏，所以必须仔细选择最优的设计点。

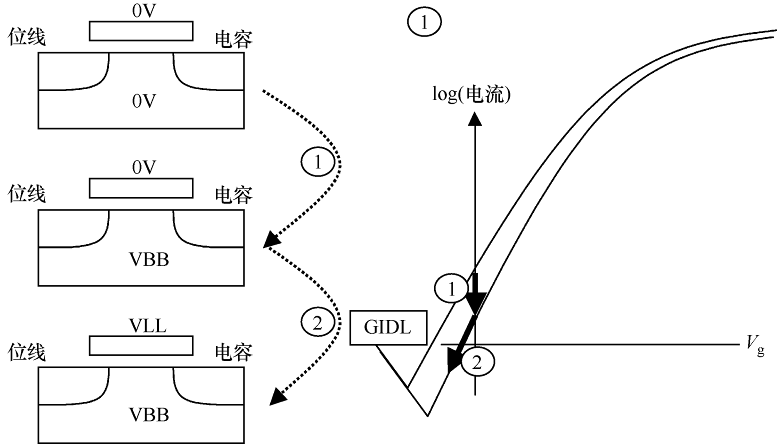


图 7-7 控制阵列器件泄漏的电压工程。图右半边是采用对数坐标表示的传送器件电流，以着重显示亚阈区电流。左半边显示了新引入电压的过程。首先，加负的 p 阱电压 VBB，改变了 *I-V* 曲线。在 100nm 节点附近，未选中器件的字线电压被设定为负值以获得较低的阈值电压，但是又不能太负，否则将进入 GIDL 区

在 100nm 工艺时代，这种相当经典的方法被证明是不够的。作为一项新的改进，引入一个新的电压 VLL[4]。作为未选中单元的字线电压 VLL，被设定为一个绝对值很小的负值，以减小阵列器件的亚阈电流。但是，必须仔细选择一个优化的 VLL 值，因为采用 VLL 虽然减小了亚阈区泄漏，但是同时会增加节点结的栅感应漏极泄漏（GIDL）。因此，开始阶段，如期望的那样，通过亚阈行为的改善使泄漏不断减小，但是随着 VLL 更负，因为 GIDL 而使泄漏从减小变成不断增大。

选用优化的 VBB、VLL 和 VAA 值，可以使阵列器件满足数据保持时间所要求的泄漏标准。最关键的器件参数是阈值电压，它必须被设定到足够大的值。然而，亚阈斜率不会随着几何尺寸缩小。这样，为了保证保持时间，在几代工艺间必须阈值保持相对不变。然而阵列器件的导通电流要求必须支持不断增加的单元读取速度，以减少存取时间。这种情况完全不同于逻辑工艺的情况。对于逻辑工艺，通过工艺缩小获得性能的改善，但是以持续增长的关态泄漏为代价。

对这个问题，最直接的解决方法是采用升高的栅电压 VPP。从 1980 年以来，已经有好几代工艺在 DRAM 中采用升高的 VPP，主要是要求将全部的 VAA 电平写入存储器单元，这样就要求传送器件的栅电压满足：VPP > （VAA + V_{th}）。就我们所讨论的关于工艺缩小问题而言，升高的 VPP 会额外增加读取条件下的驱动电流。因此，当技术边界条件仍然满足时，选择尽可能高的 VPP 很有优势。

然而，栅氧厚度又设定了一个严格的限制——氧化层中的电场不能超过击穿（GOI）限制，即不能出现介质经时击穿（TDDB），这是一个可靠性问题。虽然击穿电压大于 10mV/cm，但是应用中最大电场被限制在约 5 mV/cm。为了将这个界限提高，现代的 DRAM 工艺采用两种栅氧厚度的器件。在外围电路中，器件使用薄栅氧以优化性能，而阵列器件则采用厚氧化层以允许升高的 VPP。另外，VPP 电荷泵的设计又提出了一个不那么严格的限制。为了使电荷泵的效率高，VPP 不能设计得太高。以前，由于电压较高，升高的字线不是太重要。然而，随着将来存储器的外部电源电压快速减小时，在这就变得举足轻重了。下一代器件会工作在 1.5V 或更低的电压下，上面的描述的原理期望会使能够缩减进入 100nm 领域。随着尺寸的缩小，在某些情况下，甚至这些办法都会变得不充分了。这样就必须考虑更多的积极方法。通过浏览 DRAM 电容的历史记录可以获得一种直接的但是在技术上具有挑战性的方案。几代工艺之前，平面的二维电容被三维纵向电容取代。沿着相同的路线，必须探究进一步的革新以继续缩减能进入亚 100nm 技术。存取晶体管是下一个制约缩减的对象，需要采取创新措施，以克服对缩小的制约。晶体管被垂直折叠进沟槽单元中的实验型 DRAM 已经证明是可行的[5,6]。这种拓扑结构的优点是阵列器件的沟道长度与目标特征尺寸 f 无关。图 7-8 所示为对这一技术的说明。晶体管不再是一个平面器件，而是沿着垂直方向位于硅表面的位线接触扩散与深入到硅内部的沟槽之间。

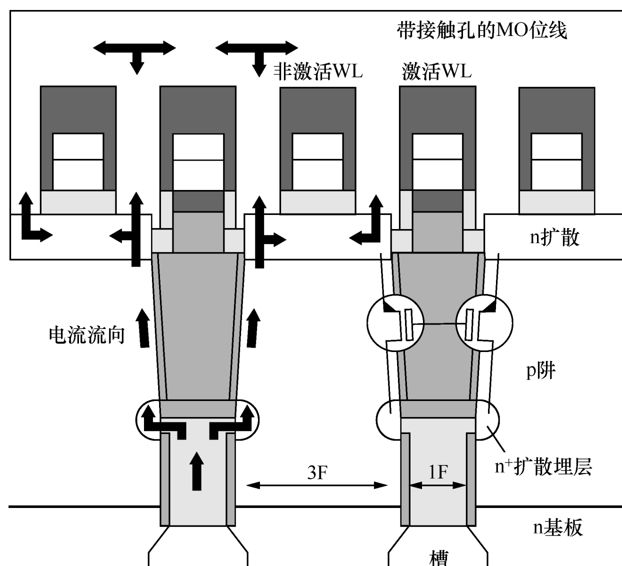


图 7-8　采用垂直存取晶体管的存储器单元

（晶体管向垂直方向延伸并实现了与存储器单元电容的三维集成。一个主要优点是存取晶体管的栅长不再与特征尺寸 f 有关）

虽然存在挑战，人们还是不断产生并实现新的想法，从而在 100nm 特征尺寸之下继续缩小 1T1C 单元。如果仔细考虑了电容和存取晶体管，并实现了所有电阻和（耦合）电容的正确建模，建立阵列将不会产生额外的严重问题。下一个挑战是读出放大器的设计。

7.5　读出放大器的缩放

读出放大器经过几代演变已经成为一个得到充分优化的电路。缩放使读出放大器产生了两个系统问题：阵列工作电压（VAA）的降低和器件面积的减小。下面先讨论电压的影响。

与存取晶体管相似，读出放大器必须满足速度和泄漏的制约条件。在读出 0 时，最关键的因素是读出速度（见图 7-9）。在单元读取之前，位线被预充电到 VAA/2。当字线打开后，由于 BLt 和存储器单元之间的电荷共享，导致 BLt 的电压下降。BLt 需要通过读出放大器的 n 型晶体管 Nt 放电，Nt 通过 BLc 控制（见图 7-3）。然而，读出之后，通过相互之间存在有间隔的相邻字线之间的线-线耦合，BLc 的电压一般会降低到 VAA/2 之下。这样，SAN 上存储 1 的读出单元会以稍稍高于存储 0 的读出单元上的电平开始。一般地，有数千个读出放大器被并行激活去读出一条字线上所有位，如果其他大多数单元存储的是 1，这时会在读出 0 时出现极大的延迟。这种情况下，针对所有位线负载读出 1 时，SAN 不能被足够快速地拉低。0 的读出因此也处于不利状态，特别是 VAA 处于很低的低电平时。如图 7-9 所示，随着 VAA 的不断减小，读出速度将快速衰退，最终由于访问速度不够快，在读取列时，扰乱了位线上的信息。由于阈值随着温度降低而增加，因此在低温时由于阈值增大，使这个效应最显著。增加读出速度的一项常用技术是降低读出器件的阈值。因此，先进的 DRAM 工艺在读出放大器中采用低阈值 n 型晶体管。一旦读出放大器晶体管处于锁存时，降低器件的阈值将增加读出放大器对的泄漏。在前几代工艺中，可以很容易地达到满足读出放大器两个制约条件（泄漏和速度）的设计点。然而，工艺缩小到 100nm 之下时，情况将发生改变，以致不再存在能同时满足两个制约条件的器件设计结果。在这个阶段，阵列电压 VAA 需要减小到 1V 的水平。这样，设定读出时过驱动级别的 VAA/2 也仅仅是 500mV。因此，为了维持足够的读出速度，需要很低阈值的器件。由于亚阈斜率的限制，在读出放大器处于锁存后，这些低阈值将导致难以管理的亚阈泄漏。

据报道已有多种方法可用来改善较低阵列电压时的 DRAM 读出性能。最简单的方法是将位线预充电到 0V[7]。在这种条件下，读出放大器的放大器件将是 p 型晶体管，而 SAP 从 0V 的预充电状态拉高到 VAA 时，这些器件将工作在完全

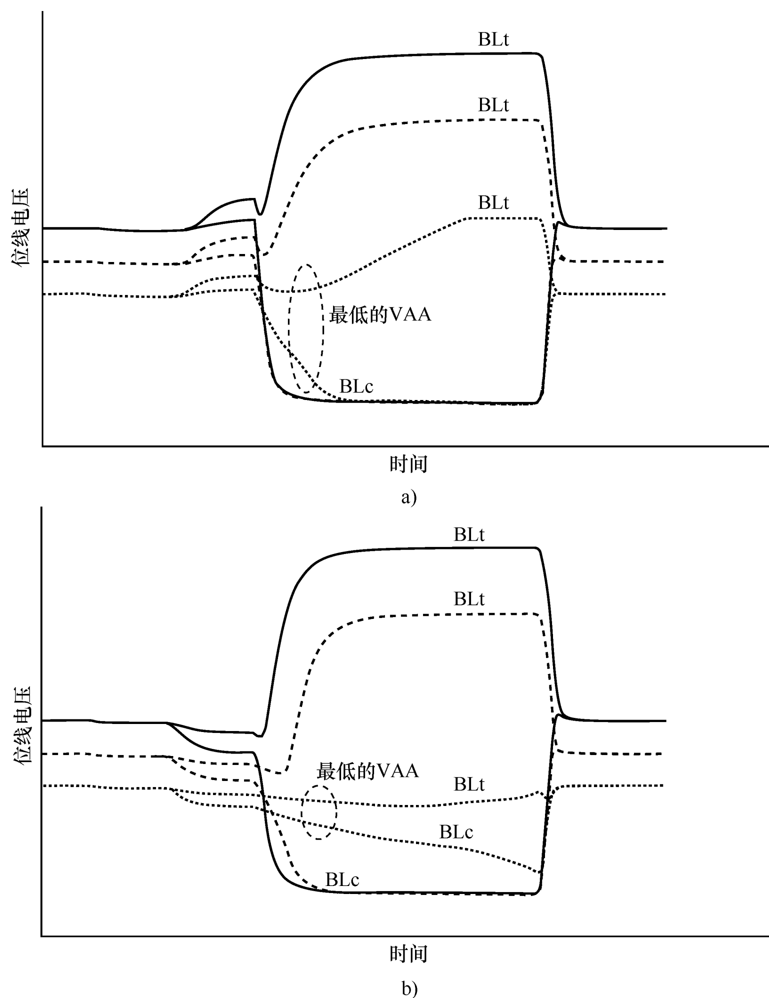


图 7-9　低温下不同 VAA 电平（实线/虚线/点线）条件下读出时的典型位线电压变化情况，随着 VAA 的降低，读出 0 时的速度迅速变差

a）读出 1 时　b）读出 0 时

的 VAA 下，因此可以轻易地实现足够的栅过驱动。虽然这种简单的方法乍看之下非常吸引人，但是它还是存在许多缺点。最显著的是需要给读出引入一个参考单元。参考电容要与存储单元有相同的电容并且必须被预充电到中位电平 VAA/2。为了读出位线上的单元，参考单元必须被连接到互补位线，反之亦然。读出一个单元将引起互补位线上电压的上升，等于全信号电压的一半（图 7-10）。这时，对存储在存储单元中的 0 和 1，可以在位线和互补位线之间获得正确的差分信号。参考单元将增加额外的面积和特殊的版图，因为它们需要一个端口给单元

充电到 VAA/2。另外，地电平预充电使阵列的工作电流加倍。在放大期间，一个位线必须从地完全拉高到 VAA。读出后，这些电荷泻放到地而消失。因此，地电平预充电是一种可行方案，但是付出的代价是增大功耗。

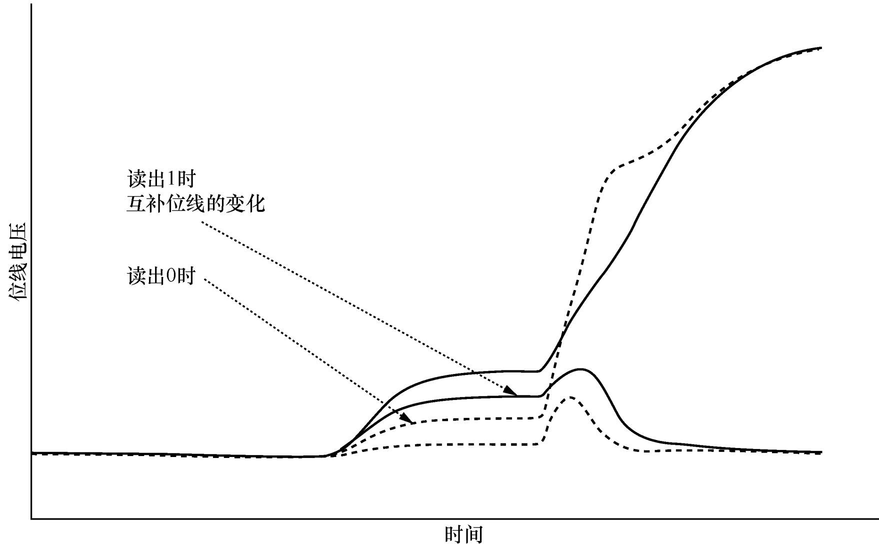


图 7-10 位线和互补位线上地电平预充电进行读出时读出电压的典型变化情况
［虚线（实线）代表读出 0（1）。需要读出互补位线上的参考单元以建立一个中点。读出位线上的 0（1）会通过电荷共享引起电压增加，其值小于（大于）参考位线。参考单元位线的差异是由线-线耦合产生的］

参考文献［8］中讨论了另一种方法。作者期望从根本上解决器件阈值随温度而变化的问题。为了补偿温度对阈值的影响，采用有源调节阱偏置的方法，在低温时降低阈值，而在高温能维持足够高的阈值。这样，可以同时实现低温时的高读出速度和高温时的低阵列泄漏。

另外一种技术是位线和读出放大器采用不同的预充电电平[9]。位线被预充电到 VAA/2，以保持中位电平预充电阵列的优势。而读出放大器被预充电到全 VAA 电平以使读出放大器器件有足够的裕量。该电路兼顾了几方面的良好功能，但是其缺点是增加了器件数并且走线比传统的读出放大器复杂。

电压减小对读出放大器工作来说显然直接施加了一个硬性制约。不断缩小的器件面积则施加一个更精细的制约。随着面积减小，由于杂质浓度随机波动引起的阈值电压的随机失配将变得重要。从 250nm 节点开始，高性能 DRAM 与模拟设计者就不断与这个问题斗争。模拟设计者可能采用的一项技术是增加器件面积，因为∂V_{th}近似地与 $\sqrt{1/\text{device area}}$ 成正比[10]。但是，这种方法在面积和节

距受限的读出放大器中是不合适的。针对这一问题，在最近的文献中提出了几种解决方法，其中使用有源偏移校正读出放大器[11-13]。这些方法中共同的原理是，在均衡过程之后，对位线的预充电电平进行前置补偿，从而抵消读出放大器的偏移。但是其带来的负面作用是，增加了器件数，读出放大器中的布线更加复杂。

7.6　总结

总之，降低的工作电压和读出放大器器件的失配对于在 100nm 之下的缩小没有明显的限制。建议提出的解决方法使 DRAM 的持续缩小变成可能。然而，为了使得具有可制造性和鲁棒性的设计得以继续缩小，还需要对于拓扑和电路设计中存在的一些问题采用新的解决办法。

参 考 文 献

[1] W. M. Regitz and J. Karp, A three transistor-cell, 1024-bit, 500 NS MOS RAM, *IEEE International Solid-State Circuits Conference*, Vol. XIII, pp. 42–43, Feb. 1970.

[2] C. N. Ahlquist, J. R. Breivogel, J. T. Koo, J. L. McCollum, W. G. Oldham, and A. L. Renninger, A 16 384-bit dynamic RAM, *IEEE J. Solid-State Circuits*, Vol. 11, pp. 570–574, Oct. 1976.

[3] K. Itoh, R. Hori, H. Masuda, Y. Kamigaki, H. Kawamoto, and H. Katto, A single 5 V 64K dynamic RAM, *IEEE International Solid-State Circuits Conference*, Vol. XXIII, pp. 228–229, Feb. 1980.

[4] H. Tanaka, M. Aoki, T. Sakata, S. Kimura, N. Sakashita, H. Hidaka, T. Tachibana, and K. Kimura, A precise on-chip voltage generator for a gigascale DRAM with a negative word-line scheme, *IEEE J. Solid-State Circuits*, Vol. 34, pp. 1084–1090, Aug. 1999.

[5] S. Wuensche, M. Jacunski, H. Streif, A. Sturm, J. Morrish, M. Roberge, M. Clark, T. Nostrand, E. Stahl, S. Lewis, J. Heath, M. Wood, T. Vogelsang, E. Thoma, J. Gabric, M. Kleiner, M. Killian, P. Poechmueller, W. Mueller, and G. Bronner, A 110 nm 512 Mb DDR DRAM with vertical transistor trench cell, *Symposium on VLSI Circuits Digest 16*, pp. 114–115, June 2002.

[6] T. Kirihata, G. Mueller, M. Clinton, S. Loeffler, B. Ji, H. Terletzki, D. Hanson, C. Hwang, G. Lehmann, D. Storaska, G. Daniel, L. Hsu, O. Weinfurtner, T. Boehler, J. Schnell, G. Frankowsky, D. Netis, J. Ross, A. Reith, O. Kiehl, and M. Wordeman, A 113 $mm^2$600 Mb/s/pin 512 Mb DDR2 SDRAM with vertically-folded bitline architecture, *IEEE International Solid-State Circuits Conference*, Vol. XLIV, pp. 382–383, Feb. 2001.

[7] J. Barth, D. Anand, J. Dreibelbis, and E. Nelson, A 300 MHz multi-banked eDRAM macro featuring GND sense, bit-line twisting and direct reference cell write, *IEEE International Solid-State Circuits Conference*, Vol. XLV, pp. 156–157, Feb. 2002.

[8] K. Hardee, F. Jones, D. Butler, M. Parris, M. Mound, H. Calendar, G. Jones,

L. Aldrich, C. Gruenschlager, M. Miyabashi, K. Taniguchi, T. Arakawa, A 0.6 V 205 MHz 19.5 ns tRC 16 Mb Embedded DRAM, *IEEE International Solid-State Circuits Conference*, Vol. XLVII, pp. 200–201, Feb. 2004.

[9] J.-Y. Sim, Y.-G. Gang, K.-N. Lim, J.-Y. Choi, S.-K. Kwak, K.-C. Chun, J.-H. Yoo, D.-I. Seo, and S.-I. Cho, Charge-transferred presensing and efficiently precharged negative word-line schemes for low-voltage DRAMs, *IEEE Symposium on VLSI Circuits*, Session 22–4, 2003.

[10] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, Matching properties of MOS transistors, *IEEE J. Solid-State Circuits*, Vol. 24, pp. 1433–1439, Oct. 1989.

[11] T. Furuyama et al., A new sense amplifier technique for VLSI dynamic RAM's, pp. 47–44, *Proceedings of the IEEE International Election Devices Meeting*, 1981.

[12] S. H. Hong, S. H. Kim, S. J. Kim, J. Wee, and J. Y. Chung, An offset cancellation bit-line sensing scheme for low-voltage DRAM applications, *IEEE International Solid-State Circuits Conference*, Vol. XLV, pp. 154–155, Feb. 2002.

[13] J.-Y. Sim, K.-W. Kwon, J.-H. Choi, S.-H. Lee, D.-M. Kim, H.-R. Hwang, K.-C. Chun, Y.-H. Seo, H.-S. Hwang, D.-I. Seo, C. Kim, and S.-I. Cho, A 1.0 V 256 Mb SDRAM with offset-compensated direct sensing and charge-recycled precharge, *IEEE International Solid-State Circuits Conference*, Vol. XLVI, Feb. 2003.

第 8 章　片上互连的信号完整性问题

8.1　引言

过去几十年中，令人注目的工艺技术的进步促使所制造的集成电路（IC）具有更小的特征尺寸、更高的集成度和更快的工作频率。虽然这些进步对提高IC的性能非常有利，但是它们也给片上互连设计带来了极大的挑战，如电感效应、串扰噪声和电源供电的稳定性等。本章主要从设计开发和性能分析的角度来说明信号完整性问题。首先，介绍高速互连的缩小趋势与品质因数。然后，讨论信号完整性分析的各种方法和设计指导，讨论的范围从互连寄生参数的提取和电路模型的表述到时序和噪声分析。最后，从物理设计优化和电路设计技术两方面，给出纳米级设计中改善信号布线质量的实用方案。

前面第 2 章已从工艺技术的角度讨论了互连缩小的趋势和挑战。图 8-1 所示为现代微电路制造中的多层片上互连结构。根据其所连接的模块规模大小，互连主要分为两类：连接晶体管和小规模电路模块的局部互连层（长度一般约为 10～500μm）；连接微处理器中较大规模功能模块的全局互连层（长度一般大于 300μm）。全局互连的典型应用包括数据总线、时钟分布、电源供电以及射频设计中的螺旋电感。而局部互连相对较短，其规模大小与工艺技术有关，因此对电路性能的影响很小，而全局互连的长度可以一直延伸到高达芯片边界的程度，并将成为纳米设计中一个重要的性能瓶颈。因此，本章重点是从设计开发和验证阶

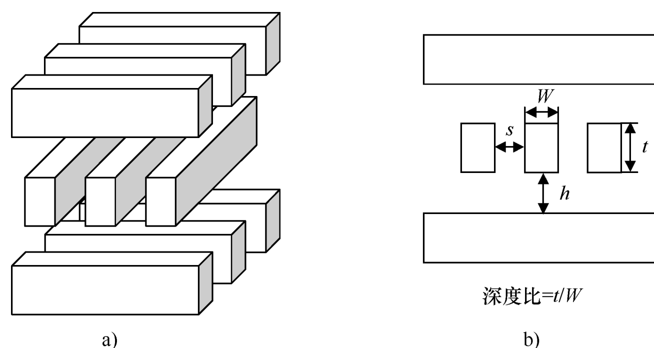


图 8-1　当前的互连结构和几何参数定义

a）三维视图　b）横截面视图

段分析全局互连的信号完整性。但是其原理对局部互连同样有效。

信号完整性是指电路工作中信号传输的质量。信号完整性问题与信号的功能（例如逻辑开关、数据传递、电源供电或时钟分布）有关，也包括了一系列的设计问题，例如串扰噪声、时延的不确定性、信号过冲、有限的压摆率、*IR* 压降，d*I*/d*t* 噪声和时钟错位与抖动等。本章将讨论高速 IC 设计中大家普遍关心的信号时序和波形失真问题。在纳米领域，信号完整性不断恶化的现象是工艺缩小不可避免的后果：为了维持不断增加的器件密度，先进工艺节点下的片上互连的节距更小，如图 8-2a 所示[1,2]。另一方面，时延减小的需求又要求金属互连横截面面积不断缩小，面积等于互连宽度 w 与厚度 t 乘积。因此，金属互连变得更窄、更高，导致了更大的纵横比（定义为 t/w），如图 8-2a 所示。这种方法在减轻了互连时延问题的同时，却在邻近的两条互连之间引起强烈的耦合。图 8-2b 所示为耦合电容（C_c）与金属对地电容（C_g）之比和互感（L_m）与自感（L_s）之比，

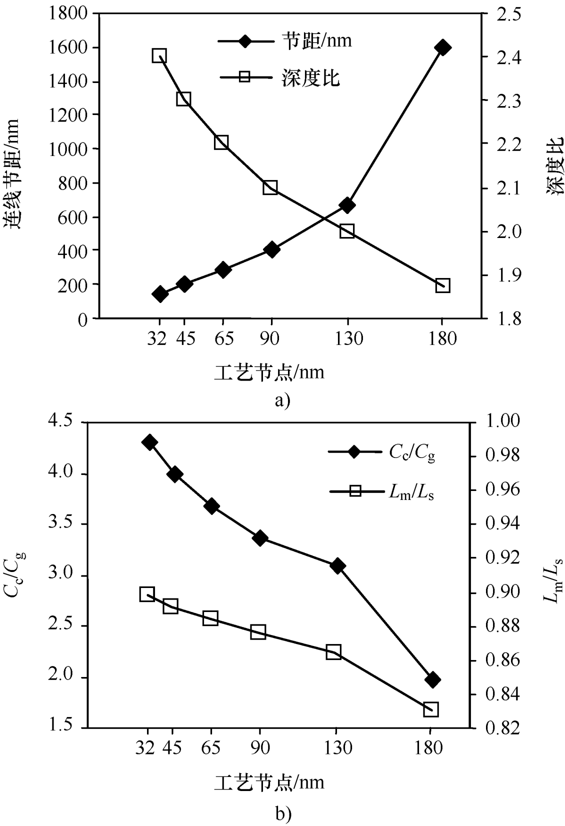


图 8-2　由于互连间距更小以及纵横比更大的结果，互连耦合随着工艺的缩小变强

a）全局互连间距和纵横比随工艺缩小的变化趋势　b）全局互连层的容性和感性耦合的变化趋势

其中假定互连间距（s）等于互连宽度。如图所示，相邻互连之间的容性和感性耦合剧烈增加：例如从 180nm 工艺节点到 65nm 工艺节点，C_c/C_g 之比将增加 70%。因此，在未来的工艺节点下，信号传输将会受到邻近互连开关的严重影响。

使串扰噪声更加严重的另一个重要因素是更快的信号开关速度。当代高性能微处理器设计已经实现了几 GHz 的片上时钟速度。因为容性耦合 $C(\mathrm{d}V/\mathrm{d}t)$ 和感性耦合 $L(\mathrm{d}I/\mathrm{d}t)$ 都是交流现象，更快的信号开关切换（即更短的 $\mathrm{d}t$）必然会引起更大量的噪声，这将是一个问题，因为电路的噪声容限随着电源电压（V_{dd}）的减小变得更差。考虑到器件的可靠性和芯片的功耗，电源电压由 180nm 工艺节点的 1.8V 减小到 65nm 工艺节点的 0.7V。同时，器件的阈值电压（V_{th}）也随着减小（虽然考虑到泄漏电流问题，与电源电压相比，阈值电压减小得相对慢一点）以保证足够的驱动能力。这就导致了数字电路（特别是动态电路）的噪声容限远低于以前，从而对鲁棒的纳米设计提出了极高的噪声控制和信号完整性解决要求。

8.1.1　互连的品质因数

在讨论解决信号完整性问题的设计方法学之前，首先引入通常被用来评价片上互连性能的品质因数（figures of merit，FOM）：

1. 信号时延和能量损耗

信号时延（时序）是同步设计中主要的品质因数指标。如图 8-3 所示，信号时延是从驱动门输入信号 50% V_{dd}位置点到互连末端（即接收器的输入）处信号 50% V_{dd}位置点之间的时间间隔。时延是驱动门和互连负载的函数。在 0.25μm 和高于 0.25μm 的工艺，主要的互连负载是金属对地的电容，因此互连负载与金属互连的长度成正比。随着工艺的缩小，互连负载的主要部分已经变成了耦合电容。因此，信号时延不仅是局部互连几何尺寸的函数，也与相邻互连的信号开关活动有关。未来设计中的另外一个问题是互连的能量消耗（$C\cdot V_{dd}^2$）。为了实现更高的集成度，芯片中互连总长度急剧增加（例如，在 45nm 工艺节点，可以长达 5km），导致了更大的总电容和更大的能耗。要注意互连能耗与总电容以及 V_{dd} 的平方成正比，并且基本上与互连电阻和电感无关。

2. 串扰噪声

一般来说，噪声源于给定互连与相邻互连通过耦合电容或者互感的耦合。在 180nm 工艺节点，交叉耦合仅仅对高性能的设计有严重影响。然而，当进入 100nm 以下时，对所有 IC 设计，互耦合就成为与时序、功耗和面积同等重要的问题之一。串扰噪声严重影响着电路的性能，主要表现在两个方面：首先较大的噪声会触发并引起错误的逻辑开关动作，并且因此会造成故障；其次也是更普遍

的问题是串扰噪声使原来的开关波形失真，并且与开关的时刻有关，会引起不同程度的严重时延（在 65nm 工艺节点会高达 80%）。这样，如果设计者继续使用当前的最坏情况设计方法，针对这种时序不确定性来预留合适的容限，结果将是无法承受的过度时序预算，而且还会使性能变得很差。在 8.3.4 节，我们将讨论对噪声敏感的时序分析。

3. 波形完整性（压摆率、过冲和信号振铃）

在数字设计中，理想的信号是一个在 0 和 V_{dd} 之间瞬时开关的阶跃函数。然而，这种模型仅是实际开关波形的近似。实际信号的压摆率不会是零，并且可能有信号过冲和振铃。压摆率（或上升边）定义为信号从最终值的 10% 变到 90%（对目前更新的设计，从 30% 到 70%）时需要的时间。由于驱动器强度有限，互连是有损耗的，实际的信号开关会有一定的压摆率，这将延缓了信号的稳定和传输时间。当电路速度快速升高时，需要对这部分的时序预算严加控制。更快的电路速度还会导致其他的高频电磁效应，如不希望的传输线行为和很难实现阻抗的匹配。另外，升高的电路速度也会导致信号过冲（即高于 V_{dd} 或低于地）和振铃，如图 8-3 所示。信号的振铃破坏了趋稳时间；过冲会引起接收器中栅氧的可靠性问题，特别是传输门电路。当工艺缩小的趋势持续下去并且数据通信的速率接近每秒太位时，目前这种互连结构中的波形失真最终会限制带宽[3,4]。因此需要新的互连工艺和设计技术来克服这种限制。

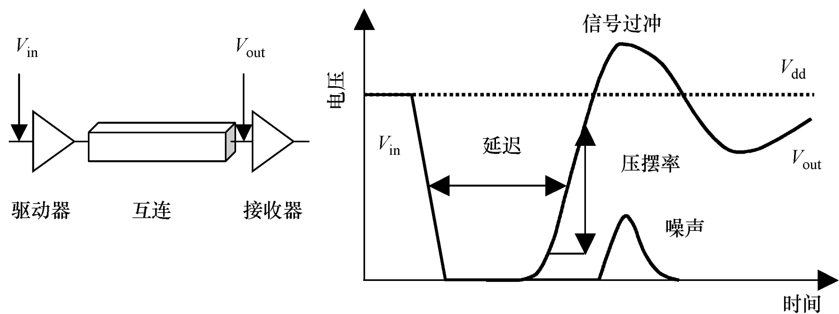


图 8-3　片上互连分析时主要的品质因数定义

从设计的角度看，需要仔细研究这些品质因素以防止信号完整性退化。这些指标将贯穿应用到本章的性能分析和设计方案中。受最近的工艺——设计联合优化的推动，这些性能品质因素还会被进一步整合进工艺方案中，以使片上通信的吞吐量最大化。

8.2　互连参数提取

为了进行时序和信号完整性分析，必须提取互连版图以及工艺信息，例如互

连的长度和宽度、邻近互连以及相关介质情况。将它们变换为电气参数表征，并与其他电路单元结合在一起，进行性能评估。基于设计和工艺规范，一条物理互连一般会被转换为由电阻、电容，如果必要的话还有电感组成的网表。常规的互连建模方式是分段采用集总 *RC* 等效。然而，随着电路工作频率的上升，在针对高性能互连建模时，这个模型缺乏有效的精度。随着工艺尺寸的缩小，在设计的前期规划和设计的后验证阶段中所采用的提取方法应该包含更多的物理型信息。例如，在以前的模型中被忽略的互连电感是影响现代设计中一些全局互连性能的重要因素。本节将讨论互连 *RC* 和 *RLC* 参数提取的原则和可行方案。另外，我们也给出了片上寄生参数的表征技术，可以用于设计校准和模型验证。

8.2.1　互连的等效电路表示

金属线基本电气行为可以用下述麦克斯韦方程组描述：

$$\nabla \cdot D=\rho \quad \nabla \times E=-\frac{\partial B}{\partial t} \quad \nabla \cdot B=0 \quad \nabla \times H=J+\frac{\partial D}{\partial t} \tag{8.2.1}$$

再结合下面的电荷守恒定律可以进行完整确切地表征：

$$\nabla \cdot J+\frac{\partial \rho}{\partial t}=0 \tag{8.2.2}$$

然而，直接求解这些方程需要实际上难以操作的大量计算。因此，根据所关心的频率范围和互连长度，通常对这些方程进行简化，以改善计算效率。图 8-4 所示为几个基于不同近似的片上互连的电路模型。

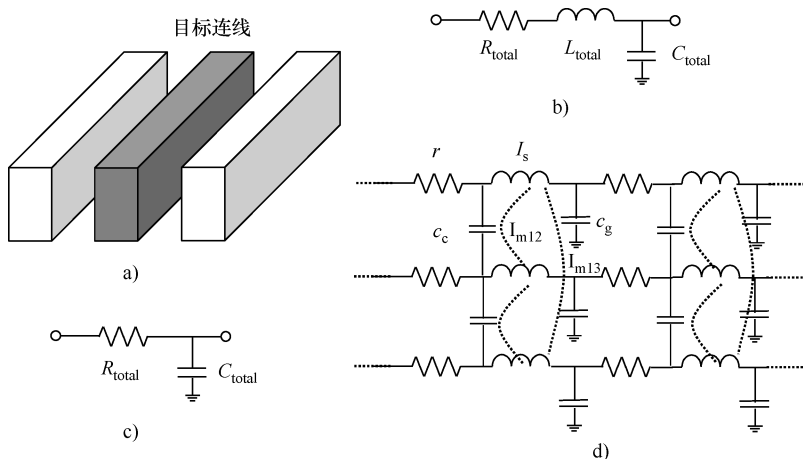


图 8-4　片上互连分析的各种电路模型

a）三条平行互连　b）集总 *RLC* 模型　c）集总 *RC* 模型　d）分布 *RLC* 模型

金属线的行为是与频率相关的。在直流情况下，它表现为电阻，会引起电源压降（IR 降）和静态功耗（IR^2）。在交流情况下，互连上的信号传输特性受电场和磁场相互作用的影响。在当前的 IC 设计中，如果信号频率相对较低并且互连长度与信号的波长相比非常短（例如 10GHz 时，在 $k=3.0$ 的介质中波长为 1.7cm），通常可以采用准静态假设。在准静态假设下，可以忽略式（8.2.1）中的位移电流$\frac{\partial D}{\partial t}$项，也就忽略了电场和磁场之间的关联。因此，可以分别确定和提取互连电容 C 和互连电感 L，最终可以采用 RC 或 RLC 等效电路描述互连，如图 8-4 所示。对一条均匀的互连，麦克斯韦尔方程可以简化为电报方程（传输线理论）来求解电气响应：

$$\frac{\partial^2 V}{\partial x^2} = RC\frac{\partial V}{\partial t} + LC\frac{\partial^2 V}{\partial t^2} \tag{8.2.3}$$

式中，x 为长度；t 为时间；V 为电压。另外，如果金属线远短于信号波长，将整个互连处理为集总模型是一种可靠的近似，因为信号电平在整个互连上近似为恒定的。图 8-4a 显示的实际互连可以分别用图 8-4b 中所示的集总 RC 或 RLC 模型代替。确定一条互连是否能用集总模型代替的经验法则是，检验它的长度是否满足下述准则：

$$\text{Length} \leqslant \frac{\lambda}{20} \tag{8.2.4}$$

式中，λ 为信号的波长。注意，一个数字信号的频谱主要取决于上升时间 t_r（如 $\frac{1}{3.14t_r}$）而不是信号频率本身。因此，λ 应该从信号的上升边来估计。例如，对一个在 $k=3.0$ 的介质中上升边为 20ps 的互连，只要其长度小于 500μm 就可以使用集总模型。全局互连的长度一般太长不能满足上面的条件，因此需要被分成相等的片段，使每个片段都满足式（8.2.4）。每个片段可以采用由简单的 RC 或 RLC 单元组成的阶梯型、T 型、或 Π 型电路。图 8-4a 所示的三根耦合互连的分布 RLC 等效如图 8-4d 所示。在进行性能分析时，集总模型便于对缩放和物理现象的理解，因此特别适合设计开发的前期阶段。然而，在设计阶段后期需要更高精度时，应该使用分布模型。

何时何处需要考虑电感 电感是对电流所产生磁场分布的唯一度量。因为在时钟频率较低时可以忽略互连电感，片上互连一般仅采用 RC 元件进行建模；电感是芯片外封装设计者考虑的问题。然而，当时钟频率进入到 GHz 领域时，互连电感形成的感抗（$j\omega L$）将可以与互连电阻 R 相比拟，甚至会成为全局互连总金属阻抗（$Z=R+j\omega L$）的主要部分。另外，低电阻率的铜和低 k 介质的引入也使 L 的重要性相对增加了。在许多重要的电路设计中电感的影响将不再可以忽略；电感不仅增加了信号时延而且也引起了电压过冲和压摆率的减小，这会进而

增加相邻互连的串扰噪声。因此，电感的影响成为信号完整性分析中日益重要的问题。另一方面，因为考虑 L 一般需要复杂的计算，所以只有必要时才将 L 考虑进去。从信号时序的角度看，可以通过估计下面这些时间常数加以判定 L 的重要性：

（1）驱动器输入信号压摆率：t_r

（2）传输线上的飞越时间 TOF：$t_f = \text{length} \cdot \sqrt{LC}$

（3）RC 互连的 Elmore 时延：$t_e = \dfrac{RC}{2} \cdot (\text{length})^2$

其中，R、L 和 C 分别为单位长度的电阻、电感和电容。考虑到品质因数，如果互连满足下面两个条件[5,6]，互连应该被建模为 RLC 互连：

（1）$t_f > t_e$。在这个条件下，电感的影响导致较大的信号时延。RLC 模型表示的是传输线，比较 RLC 和 RC 模型的时间常数，相当于评估传输线的阻抗（$Z = \sqrt{L/C}$）与互连的总电阻 $R \cdot \text{length}$。当 Z 比总电阻大得多时，RLC 的影响对估计互连性能将非常关键。

（2）$t_r < 2t_f$。因为 $2t_f$ 是信号从驱动器到互连末端穿越整个来回行程所需的时间，这个条件意味着如果开关足够快，信号传输会受到反射波的影响，表现出传输线特性。结合互连长度和 RLC 单元之间的关系，可以将这两个条件转化总结为

$$\frac{t_r}{2\sqrt{LC}} < \text{length} < \frac{2}{R}\sqrt{\frac{L}{C}} \tag{8.2.4a}$$

如果左边的条件大于右边的条件，即

$$t_r > 4L/R（输入信号不够快） \tag{8.2.5}$$

在这种情况下，不管互连有多长，电感效应都可以忽略。图 8-5 表示了在什么范围内 RLC 对一条全局信号互连非常关键的一个实例。电感的效应对中间层互连（如典型的 130nm 工艺，时钟频率为 1.2GHz，互连长度约为 2～10mm）更加明显。在先进工艺中大多数互连都在这个长度范围内，所以需要表示为 RLC 互连。局部金属互连和中等范围的金属互连可以建模为 RC 电路，因为它们主要表现为电阻性。图8-5表明随着频率的升高，越来越短的互连都将呈现出电感效应。需要说明的是，上面的讨论是基于单根互连。实际上，在同一层会有很多相邻的互连，因

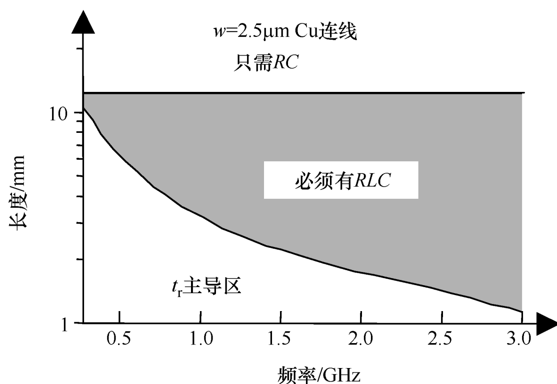


图 8-5　电感显著区域实例

此式（8.2.4a）和式（8.2.5）中的电感应该包括电流返回回路的总电感。传输线行为除了受互连布图的影响之外，还与互连两个末端的边界条件有关，通过调节驱动器电阻和负载电容可以改变边界条件。

当信号频率持续升高到 10GHz 以上并且互连长度持续增加到超过信号波长时，准静态的假设就不再有效。位移电流的出现，使得各单元之间的相互作用不再是瞬时的，从而信号退化成为正确分析的关键。因此，互连参数的提取需要与电气分析结合起来才能求解出时域和频域的响应。虽然这不是当前和近期设计必须考虑的问题，但是如参考文献［7］中所述，互连分析最终将必须采用全波方式。

8.2.2 *RC* 的提取

虽然电感效应对片上互连的建模日益重要，但是 *RC* 等效电路对片上大多数互连建模而言仍然足够精确，特别是局部层和中间层互连。这种参数提取对设计前期和版图后期而言都是必需的。在早期的自顶向下设计流程中，为了便于设计综合并发现主要的时序及信号完整性问题，将互连寄生参数考虑进来是至关重要的。当分层设计最终转化为平面设计后，需要从版图提取互连寄生参数值以验证设计要求是否得到满足。一般情况下，对一个定义明确的版图图形，应用三维电磁场求解技术可以获得 *RC* 和 *L* 值。在实际的设计周期中，这些数值方法（如有限差分法、有限元法、矩量法）可以实现非常高的精度，但是其面临着两个重要的限制。首先，在前期的设计规划阶段，不可能为这些场求解器提供详细的版图信息，因此这种方式的灵活性就受到限制。其次，当一个芯片上晶体管的总数超过几百万个时，全电容矩阵的计算极其耗时，实际上很难采用。由于这些原因，人们在设计时采用解析模型或查找表方法，并结合版图模式识别算法来获得运行高效的参数提取。这些解析模型或查找表模型的精度与需要花费很大代价从三维场求解器或测试结构测量结果中获得的数据相比并不逊色。

***RC* 提取的建模方法** 片上互连结构一般由矩形横截面的金属互连构成，在各层间使用曼哈顿布图方式（见图 8-1）。这种结构不仅简化了制造过程和布线算法，而且极大地减小了 *RC* 建模工作的复杂度。对宽度为 w、厚度为 t 的金属（见图 8-1b），可采用下式计算其单位长度的直流电阻 R：

$$R = \frac{\rho}{wt} \tag{8.2.6}$$

式中，ρ 为金属电阻率（铜：$\rho = 2.2\mu\Omega \cdot \mathrm{cm}$，铝：$\rho = 3.3\mu\Omega \cdot \mathrm{cm}$）。例如，典型的 $w = 0.8\mu\mathrm{m}$ 和 $t = 0.8\mu\mathrm{m}$ 的 3mm 长全局互连，在直流条件下可以建模为 103Ω 的电阻。除了金属互连，垂直连接多层金属的通孔也对路径电阻有贡献。在 90nm 工艺节点，单位面积的通孔电阻大约为 $10^{-9}\mu\Omega \cdot \mathrm{cm}^2$[1]。因此 $0.25\mu\mathrm{m} \times 0.25\mu\mathrm{m}$ 的通孔可以建模为 1.6Ω 的等效电阻。随着金属层数的稳步增加和通孔

尺寸的不断缩小，不能再忽略通孔电阻对时序模型的影响；它会给关键路径总时延带来额外 10% 的时延[8]。

金属电容描述了互连之间经由电场的耦合程度。根据耦合互连是否接地，经常将其称之为金属——地的电容 C_g（若耦合互连交流接地）；或金属——金属电容 C_c（若耦合互连是信号线）。C_g和 C_c剖面图的实例如图 8-6 所示。根据麦克斯韦方程，电场可以被金属互连屏蔽。因此容性耦合是一种短距离效应：当同一层有多条互连时，容性耦合随着互连邻近程度的增加而迅速减弱。例如，在与另一条或者更远的互连之间的 C_c（也就是在耦合互连之间至少有一条互连）与最近互连之间 C_c相比，一般减少 10%。为了简化建模并且同时维持足够的分析精度，仅需要提取并分析最直接相临互连之间的 C_c；与更远互连之间的 C_c值可以忽略。这样就形成了下面的矩阵，一般被用于共面互连（见图 8-6）的 RC 分析：

$$\boldsymbol{R} = \begin{vmatrix} \ddots & 0 & 0 & 0 & 0 \\ 0 & r_{i-1} & 0 & 0 & 0 \\ 0 & 0 & r_i & 0 & 0 \\ 0 & 0 & 0 & r_{i+1} & 0 \\ 0 & 0 & 0 & 0 & \ddots \end{vmatrix}$$

$$\boldsymbol{C} = \begin{vmatrix} \cdots & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & 0 & 0 \\ 0 & -C_{ci,i-1} & C_{gi} + C_{ci,i-1} & -C_{ci,i+1} & 0 \\ 0 & 0 & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \cdots \end{vmatrix} \qquad (8.2.7)$$

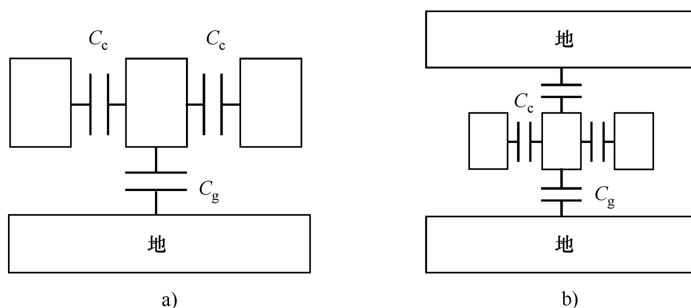


图 8-6　局部和全局互连的二维电容模型（横截面图）

a）顶层互连　b）局部层互连

每个 r_i值可由式（8.2.6）计算获得，而 C_g和 C_c元素可以由解析模型或使用如下面描述的查找表方式产生。

为了有效计算设计综合和版图验证阶段的电容，基于版图模式，多层三维互连经常被简化为二维[9-11]或准三维结构[12]。如果互连上方或下方层中的布线很密集，它们会近似为一个接地等位面，这就形成了图 8-6 所示的二维模型。这时，C_g和 C_c变成了直接与互连横截面面积相关的函数。C_g和 C_c解析模型如下所示[11]。

对顶层互连（一个接地等位面之上的金属互连，见图 8-6a）有

$$\frac{C_g}{\varepsilon}=\frac{w}{h}+2.217\left(\frac{s}{s+0.702h}\right)^{3.193}+1.171\left(\frac{s}{s+1.510h}\right)^{0.7642}\left(\frac{t}{t+4.532h}\right)^{0.1204}$$

$$\frac{C_c}{\varepsilon}=1.144\,\frac{t}{s}\left(\frac{h}{h+2.059s}\right)^{0.0944}+0.7428\left(\frac{w}{w+1.592s}\right)^{1.144}+ \qquad (8.2.8)$$

$$1.158\left(\frac{w}{w+1.874s}\right)^{0.1612}\left(\frac{h}{h+0.9801s}\right)^{1.179}$$

对局部层互连（两个接地等位面之间金属互连，见图 8-6b 有

$$\frac{C_g}{\varepsilon}=\left(\frac{w}{h_1}+\frac{w}{h_2}\right)+2.04\left(\frac{t}{t+4.5311h_1}\right)^{0.071}\left(\frac{s}{s+0.5355h_1}\right)^{1.773}+$$

$$2.04\left(\frac{t}{t+4.5311h_2}\right)^{0.071}\left(\frac{s}{s+0.5355h_2}\right)^{1.773}$$

$$\frac{C_c}{\varepsilon}=1.4116\,\frac{t}{s}\exp\left(-\frac{2s}{s+8.014h_1}-\frac{2s}{s+8.014h_2}\right)+1.1852\left(\frac{w}{w+0.3078s}\right)^{0.25724}$$

$$\left[\left(\frac{h_1}{h_1+8.961s}\right)^{0.7571}+\left(\frac{h_{21}}{h_2+8.961s}\right)^{0.7571}\right]\exp\left(-\frac{2s}{s+3h_1+3h_2}\right) \qquad (8.2.9)$$

式中,ε 为介电常数,尺度变量的含义如图 8-1 所示(式(8.2.9)中的 h_1和 h_2分别是金属线上方和下方的介质厚度)。这些模型表达式是基于物理分析,其中的系数值是通过对场求解结果拟合得到的。因此,它们可以很容易地进行缩放,并可以实现误差仅为 5% ~10% 的精度。获得单位长度 C_g和 C_c的二维值后,再乘以互连长度就得到总的电容。

在更一般的情况下,长互连可以首先被分成几个片段,每个片段可以与预先定义的版图模式匹配。版图模式的选定取决于同一层以及上面一层或下面一层的布线情况[12,13]。采用已经通过场求解器或半导体测量验证过的解析模型或查找表,求出每个片段 C_g和 C_c[13],就可以构造出 C 矩阵。然后,再与 R 矩阵一起,就可以进一步使用分析工具求解 RC 时序和噪声特性。

片上寄生参数的表征　寄生参数表征技术,不仅用于模型验证,它对于从半导体测量直接建模甚至更重要。例如,如果有简单适用的技术可以测量电容值,典型版图模式的电容值可以直接从测试结构中提取,并用于建立查找表,从而减小了过程的不确定性和模型误差。因为与片外金属互连相比,片上互连更窄而且更短,所

以他们具有更小的电容（小于 100fF/mm）和电感（小于 1nH），因此，要有一种易实现的高分辨率表征技术。*RLC* 和电气参数（电压和电流）之间的关系为

$$R=\frac{V}{I}\quad I=C\frac{\mathrm{d}V}{\mathrm{d}t}\quad V=L\frac{\mathrm{d}I}{\mathrm{d}t} \tag{8.2.10}$$

根据第一个方程，金属电阻值可以容易的从 *I-V* 测量获得。片外电容可以通过 *LCR* 仪器测量表征。人们更希望能实现片上互连电容的提取，由于它们的值较小。根据方程（8.2.10），基于电荷的电容测量（CBCM）技术已经研发出来并且实现了小于 0.1fF 的分辨率[14]。这个技术的精度受限于测试电路中的晶体管失配。当电路频率高于 GHz 时，必须要有在频域（例如 S 参数表征）和时域对高频阻抗的表征技术[15,16]。这些技术通常可以同时应用于片上和片外 *RLC* 提取。然而，测试过程要比一般的本地测试更复杂，并且需要特殊的测试结构。

8.2.3　电感提取

在纳米领域，电感效应变得非常重要，特别是对全局互连。在进行 *RC* 提取时，仅需考虑互连本身和离它最近的相邻互连。电感耦合则与此不一样，它是长距离效应；又因为磁场很容易穿透金属表面，随着互连间距的增加衰减很慢。电感的基本定义为

$$L=\frac{\oint_A B\mathrm{d}s}{I} \tag{8.2.11}$$

式中，I 为电流；B 为 I 感应出的磁场；s 为积分回路。如果 s 与 I 的路径相同，L 就是自感；如果 s 路径与电流路径不同，式（8.2.11）定义的就是互感。这个定义表明电感具有沿着回路进行计算的特点，因此求解电感性特性必须考虑整个电流回路。然而，现代互连结构中（如图 8-3a 所示的驱动器-互连-负载电容），没有直流路径构成意义明确的回路。结果，返回电流一般要流经很长的范围，使得分析非常麻烦。因此，在进行提取分析时应考虑所有可能进入电流回路的所有相邻互连。

局部和回路电感　因为返回电流路径的不确定性，实际设计中很难计算回路电感（由式（8.2.11）定义）。为了解决这个困难，人们引入了局部电感的概念，其中假定感应电流从无穷远处返回，这样就回避了定义返回回路的问题。这种称为局部元件等效电路（Partial Element Equivalent Circuit，PEEC）法的电感计算技术非常适合设计自动化，因为它仅与互连的几何尺寸有关。基于 PEEC 法，一个由每个互连片段计算得到的 L 值所组成的通用局部电感矩阵如下所示：

$$\boldsymbol{L}=\begin{vmatrix}\cdots & \cdots & \cdots & \cdots\\ \cdots & L_{si} & L_{mij} & \cdots\\ \cdots & L_{mji} & L_{sj} & \cdots\\ \cdots & \cdots & \cdots & \cdots\end{vmatrix} \tag{8.2.12}$$

然后，这个矩阵再与 RC 矩阵［式（8.2.7）］一起，借助于电路仿真器的计算，其结果就可以用于求解电流回路。这里，每个电感元件（即式（8.2.12）中的自感 L_s或互感 L_m）都可以使用场求解器，如 FastHenry 和 Raphal，或解析解的形式计算。对矩形截面的互连，解析解首先由 Rosa 和 Grover 得到[18]，且当 $l>>w$、t 和 d 时，简化为下面的关系式[19]：

$$L_s=\frac{\mu_0}{2\pi}\left[l\ln\frac{3l}{w+t}+\frac{l}{2}+0.2235(w+t)\right] \tag{8.2.13}$$

$$L_m=\frac{\mu_0}{2\pi}\left(l\ln\frac{2l}{d}-1+d\right)$$

式中，μ_0为介质的导磁率；w、t 和 l 分别为片段的宽度、厚度和长度；d 为两条互连中心线之间的间距；L_m是两个长度相等互连的互感（在参考文献［19］中也给出了不相等互连 L_m更普遍解的形式）。这些表达式说明了电感与片段长度之间是非线性关系。因此，与 RC 提取时其大小随互连线的长度缩放相比，L 的计算必须考虑整个互连的长度。而且，式（8.2.13）中的对数函数意味着，同 R 和 C 与互连几何尺寸的相关度相比，L 与互连几何尺寸的相关度更弱。注意，仅位于同一层且彼此平行的互连才对感性耦合有贡献；位于相邻层上的互连并不会影响耦合，由于它们的版图走向是正交的。

虽然采用 PEEC 法进行一般的电感提取时不需要预先给出电流返回回路的信息，但是，电感矩阵的非稀疏性（由长距感性耦合引起的）导致进行深入分析的巨大计算量[20]。不像只保留短距耦合值就足够了的 $\boldsymbol{C}$ 矩阵，$\boldsymbol{L}$ 矩阵不能通过截断加以简化；简单地舍去较远相邻互连间的 L_m值会引起模型的不稳定[21]。为了改善这种复杂矩阵的计算效率，已经做出了大量的努力。一个实例就是 $\boldsymbol{L}$ 矩阵截断法，它使用电源网格作为提取电纳矩阵的边界，而电纳矩阵是 $\boldsymbol{L}$ 矩阵的逆矩阵并且具有所需要的稀疏性[22,23]。虽然迄今为止已经开发出了大量有效的计算技术，但是仍然没有一个与 SPICE 兼容且对片上互连结构而言既简单又通用的求解方法。一个近似的回路电感模型对评估电感的物理含义（式（8.2.11））是很有用的，特别是在设计开发的前期阶段。这个模型可以用查找表[24]或解析模型的形式描述。后一种形式特别适合于与相邻互连之间通过电源和地线加以屏蔽的专用全局时钟结构[25,26]。

频率相关的 $\boldsymbol{R(f)L(f)}$ 以前只是在封装和微波设计时，由于互连线的尺寸太大，才考虑 R 和 L 的频率相关性。然而，当芯片的工作频率上升到 GHz 领域

时，片上的互连也需要考虑这些效应。这是因为高频时电流穿通金属的深度（趋肤深度）变得可以与全局互连的横截面尺寸比拟甚至更小。例如，1GHz 时，铜的趋肤深度约为 2μm；随着频率的上升，趋肤深度以频率平方根的速率减小。结果，金属互连上的传导电流密度不再均匀，并且金属阻抗变得与工作频率有关。由于采用常数 *R*、*L*（直流时的 *RL*）表示片上互连的通常做法没有考虑这种频率相关性，因此不再适用。图 8-7a 所示为使用 Raphal 法（一种基于 *RLC* 的提取工具）对三条平行互连的横截面电流分布的分析结果。当频率升高时，电流流向互连的表面并远离电流同向的邻近互连。这种非均匀型电流分布在单互连时被称为趋肤效应，而在邻近互连存在感性耦合时为邻近效应。在频率较高时这些效应导致很大的电阻，如图 8-7b 所示（注意，互连电感仅轻微下降并最终饱和）。

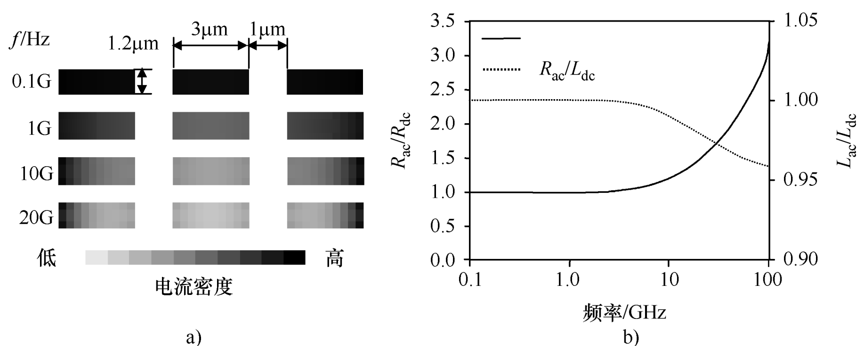


图 8-7　$R(f)$ 和 $L(f)$ 的 PEEC 仿真结果

a）横截面电流密度分布（铜互连厚度 = 1.2μm）　b）*R* 和 *L* 的频率相关情况

已有文献报道了很多分析 *R* 和 *L* 与频率相关效应的方法：采用高频 *RL* 值[26]；预先确定回路 *RL* 值[25]；或者等效电路解析模型[27]等，对 GHz 领域的时序进行估计，增大了提取和分析的复杂度。而事实上，采用直流 *RL* 和采用 $R(f)L(f)$ 之间的波形对比显示，所预测出的时延差异却非常小：图 8-8a 所示为对 90nm CMOS 工艺[2]情况下的预测，使用直流 *RL* 模型与使用 $R(f)L(f)$ 模型[28]所得到时延和上升边非常一致。这个现象的原因可以解释为在对应上升边的电压响应中，感抗 ωL 占主导：在当前的铜互连工艺下，当开关频率超过几 GHz，趋肤效应明显，ωL 通常远大于 *R*。因此，在 GHz 领域，时延对 *L* 的变化比对 *R* 的变化更敏感。当频率进一步升高时，*L* 仅轻微下降（见图 8-7b）；并且 *R* 和 *L* 与频率的关系相反，又将进一步减小 $R(f)L(f)$ 对信号时延的整体影响。总之，采用直流 *RL* 值对时延进行分析就已经足够了。上升边过后，输出信号慢下来而且电阻在波形的过冲和振铃中起主导作用，造成两个波形具有不同的振铃幅度和周期。

与时延对 $R(f)L(f)$ 的不敏感性相反，由于高频时电阻值更大，电源线上的

图 8-8　$R(f)$ 和 $L(f)$ 对电路性能的影响

a）采用斜坡输入（$w=5\mu m$，$s=2\mu m$，长度 $=3mm$；$w_{Vdd}=10\mu m$）的输出波形

b）地线上的噪声（对电源线，$w=50\mu m$，$length=500\mu m$；$C_{decoupling}=50pF$；$I=100mA$）

$L(di/dt)$ 噪声被 $R(f)L(f)$ 强烈地抑制。如图 8-8b 所示，使用 $R(f)L(f)$ 预测得到更小的峰值和更快的衰减。这就意味着，当考虑金属阻抗的频率相关性时，使电源供电稳定所需的去耦电容更小。对图示实例，为了将峰值噪声减小到 V_{dd} 的 10% 以下，直流 RL 预测需要 134pF 的去耦电容，而 $R(f)L(f)$ 的预测值为 115pF。考虑了频率相关性，采用更小的电容可以节省超过 15% 的芯片面积。这一效应很重要，因为对电源稳定性的要求越来越高导致了去耦电容器面积成本的快速增加。因此，在数 GHz 领域，正确考虑 $R(f)L(f)$ 有助于减缓这种担心，并提供有效的供电稳定性。

片上和片外电感的差异　在过去的几十年间，电感一直是板级封装这类片外互连设计中所关注的焦点。尽管对片外设计已经做了大量的关于电感建模方面的工作，但是这些工作不能直接被片上互连采纳，因为片上互连有更复杂的互连环境和不同的几何尺寸。片上和片外互连之间的主要差异可归结如下：

（1）电流返回路径。在片外设计中，接地等位面一般在版图中被设计得很宽大以减小电感（例如，带状线结构）；这些额外的互连并没有给设计增加太大的开销。最终的电流返回路径定义是明确的，因此很容易得到分析电感用的近似公式。相比之下，由于布线资源有限，片上互连一般没有定义明确的电流返回路径。

（2）电阻性损耗。片外互连比片上互连有更大的横截面。它们的损耗小于片上互连并且因此表现出更显著的传输线行为，如波的反射。低损耗传输线理论可以被应用到片外互连分析。然而，片上传输线经常表现为高损耗，从而分析起来更复杂。

（3）布线的复杂性。片上互连布线比片外互连更密集，并且需要将非常大量的邻近互连包括到分析中，才能正确地估计电流返回路径和互连行为。

（4）端接。使用电阻器与互连的特性阻抗[$Z_0=(L/C)^{1/2}$]匹配，就比较容易实现片外互连的端接。相比之下，实现理想的片上互连端接极具挑战性，因为片上互连的特性阻抗不是纯电阻性的($Z_0=[(R+\mathrm{j}\omega L)/\mathrm{j}\,\overline{\omega}C]^{1/2}$)，并且驱动器的尺寸一般是针对时延最小化要求优化过的，它的输出阻抗不可能等于 Z_0，并且片上负载的输入阻抗几乎纯粹是容性的。

8.3 信号完整性分析

在将物理版图信息转化为等效的 *RC* 或 *RLC* 元件后，按照频率范围及精度要求的不同，可以用通常的电路仿真器（如 HSPICE，SPECTRE），或者用解析模型法分析互连性能。在常规的性能驱动设计中，信号路径时延是设计开发和优化中关注的唯一焦点。然而，由于工艺尺寸的快速缩小和日益紧缩的时序预算，以前被忽视的信号完整性问题，包括串扰噪声、信号压摆率和电压过冲，成为以互连为中心的设计所要考虑的核心问题。正确而充分地考虑这些问题，对成功的芯片设计和实现至关重要。人们已经认识到这一点并用于 130nm 以及更先进工艺的许多设计案例中。特别是时序分析工具一定要能够求出信号完整性退化的情况，因为其幅度已经变得可以与时序标称值相比拟。本节首先给出一种有效实用的技术来分析物理设计中 *RC* 和 *RLC* 互连的信号完整性问题。为了在前期的时序分析中（如，全局布线和综合阶段）同时考虑串扰噪声的影响，从而实现时序的快速收敛，需要深入讨论在时序分析中包含有噪声的设计方法学。

8.3.1 互连驱动器模型

为了简化设计，片上互连的信号传输可以方便地分为两部分（见图 8-9a)：从门（驱动器）的输入到门的输出，以及从互连的近端到远端（即接收器的输入)。这样信号时延被分解为门时延和互连时延。首先分别分析每一部分，然后将结果合并在一起计算总的时序。由于局部电路互连较短，因此局部电路的性能一般由门时延决定。然而，对于全局信号，互连时延至少与门时延一样重要，甚至有时由于互连过长，互连时延成为信号总时延的主要成分。为了减轻这个效应，全局互连驱动器的尺寸，包括逻辑门和插入的转发器都应该优化，使总的路径时延最小。另外，即使驱动器没有产生多余的噪声，其尺寸也强烈影响着串扰噪声的幅度：处于静态时的大驱动器提供了更好的直流到地连接，因此能够同时较好地抑制容性和感性耦合噪声。为此，获得适当的驱动器和门负载模型对有效分析和优化都非常重要。

一个开关驱动器可以建模为一个时变电压源[29]或电流源[30]，分别如图 8-9b 和图 8-9c 所示（注意接收器被简单地建模为在互连远端的负载电容器）。由一

个斜坡电压源和一个线性电阻 R_{dr}组成的 Thevenin 等效模型（见图 8-9b）描述了门与互连负载之间的相互作用情况。例如，在典型的门时延分析中，使用模型降阶技术（如基于矩匹配的渐进波形估计[31-33]），*RC* 或 *RLC* 互连一般被近似为一个有效电容 C_{eff}或单 Π 型电路（分别如图 8-9c 和图 8-9d 所示）。在这一近似下，计算 R_{dr} C_{eff}乘积可以很容易地求得门时延。另一方面，如果将开关门建模为单个电阻器，将很难准确地预测出门输出信号的实际压摆率，特别是如果输入的压摆率和负载电容的变化范围较大时更是如此。为了克服这一缺点，在实际的 *RC* 分析中，R_{dr}和 C_{eff}值通过多次迭代匹配门输出波形上两点（如 50% 和 90%）的方法拟合确定。例如，在表征单元库元件时，上述方法产生了一个作为负载电容和输入压摆率函数的 R_{dr}值查找表。除了可用于预测门时延之外，Thevenin 模型也提供了优化驱动器尺寸以使路径时延最小化的基础。在 *RC* 分析中采用上述做法的经验法则是下述条件成立[34]：

$$\text{门时延} = \text{线时延} \tag{8.3.1}$$

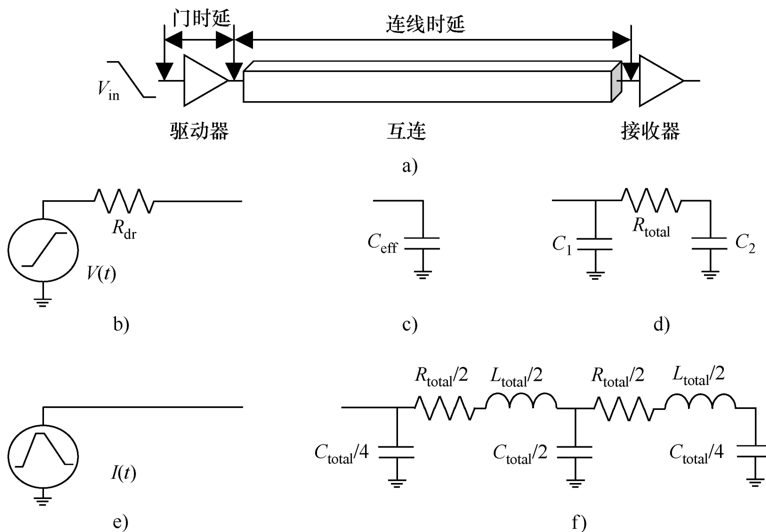


图 8-9 计算门时延采用的各种驱动器［b）、e）］和互连负载［c）、d）、f）］模型

a）门时延和互连时延的定义 b）Thevenin 模型 c）有效电容

d）单 Π 型模型 e）Norton 模型 f）*RLC* 互连的双 Π 型模型

除了用于常规的 *RC* 时序分析之外，Thevenin 模型中的 R_{dr}值不适用于预测串扰噪声，因为静态门总是处于线性工作条件下，而开关门则工作在线性区和饱和区。应该采用一个比 R_{dr}值小的电阻对静态驱动器建模，进行信号完整性分析。而且单个线性电阻太简单，无法预测 *RLC* 分析中的全波特性；也无法获知电压

过冲和信号振铃等高阶现象。采用时变电流源模型（见图 8-9e）可以进行更精确的波形完整性分析，时变电流源模型实际是描述了整个开关范围内的晶体管行为[30]。对 *RLC* 互连，使用更复杂的等效负载模型也是必要的。例如，为了反映互连近端和远端的波形反射[30]，可以采用对称 2-Π 型 *RLC* 电路（见图 8-9f）。（注意，在单 Π 型模型中，由于电阻的屏蔽效应，C_1 和 C_2 不相等）

8.3.2　*RC* 互连分析

通过互连性能分析得到的模型可能是一个由具有不同 *RC*（或 *RLC*）元件值的分支片段构成的互连树，还有负载电容和相邻的耦合情况，但是没有悬空电容和电阻回路。这种线性系统的一般求解方法源于多种适用于时域或频域的数值技术。例如，有一种技术是将 *RC*（或 *RLC*）矩阵（分别是式（8.2.7）和式（8.2.12））与基尔霍夫电压定律和电流定律（例如式（8.2.10））方程联立，并使用矩阵逼近技术[27]求解输出电压。另一种方法是利用从输入到输出的传递函数并通过匹配多个矩的方法预测信号时延和输出波形[27,31]。这些方法不仅可以处理各种版图结构和信号开关样式，而且也能给设计验证提供非常精确的时序和噪声信息。然而，它们的作用在布局布线阶段则非常有限，因为数值技术很难将互连性能与物理版图直接相关联。另外，为了实现高精度，这些数值技术一般都要花费大量的计算时间，这就限制了它们在全芯片分析中的应用。与数值求解方法不同，解析型性能度量具有模型非常简单又可缩放的特点，适合用于设计的优化；但必须在通用性和模型精度之间做出折衷选择。为了透彻理解信号完整性问题并进一步探讨电路和物理设计技术，本节重点讨论解析模型。

在片上互连的局部层和中间层，电阻性和电容性的影响在互连对电压开关的响应中占主导地位，虽然对一些全局互连还必须考虑电感效应。在满足精度要求的前提下，由于短距离的容性耦合特性，人们更愿意采用简单有效的 *RC* 分析而非 *RLC* 分析。因此，在进行时序和信号完整性分析之前，通常启动一个基于与 8.2.1 节描述相似的甄别程序，判定是否限用 *RLC* 模型。即使工艺尺寸缩小得很快，*RC* 分析仍然具有优势，因此在大多数互连时序和串扰噪声估计中被普遍使用。

***RC* 互连时序分析**　人们已经做了大量的努力，试图对 *RC* 互连采用解析型时序度量，因为解析型时序度量能方便地建立互连性能和物理版图尺寸（如互连宽度、互连长度、间距）之间的关联。最常用的度量当属描述脉冲响应一阶矩的 Elmore 时延，它适合于各个级别的 *RC* 树分析[35]。如参考文献［36］中证明的，简单的 Elmore 时延给出实际斜坡输入时 *RC* 树 50% V_{dd}实际时延的上限，因此作为 *RC* 时延的估计是比较安全的。为了进一步改善 Elmore 时延度量的精度，并且进一步扩展到能给出对更多开关特性（例如压摆率）的预测，可以通

过渐进匹配传递函数中高阶矩的方法求出单根 *RC* 互连的全输出波形的解析解[37]。这些解析度量的精度与数值分析结果相比，偏差一般在 10% 之内，这对设计的早期阶段而言已经是足够了。但是，上述度量只能处理单互连情况或互连树情况，而不能考虑邻近互连开关的影响。然而，随着工艺尺寸的不断缩小，邻近线的耦合已经是一个日益重要的问题。

由于存在邻近互连，这些邻近互连的电气行为通过 C_c 耦合到目标互连上（见图 8-10），这就使互连的时序分析变得复杂了。为了简化这种耦合情况，可以首先将目标互连去耦为等效的单个互连，然后采用解析度量（例如 Elmore 时延）计算时序。在这种方法中，使用开关因子（Switching Factors，SF）的概念将 C_c 转换为一个有效的互连到地电容，然后与 C_g 合并以分解一对 *RC* 互连，如图 8-10 所示。开关因子的概念是基于耦合电容 C_c 上的密勒效应。这种效应可以通过考察下面的场景加以理解。如果邻近互连（即图 8-10 中的互连 B）处于静态，C_c 上的电压摆幅是 V_{dd}；然而，当 C_c 的两个端点上的电压（即 V_A 和 V_B）瞬时开关时，C_c 会经历不同的电压摆幅。这种情况下，C_c 可以近似为只有一个开关端点的到地电容，有效的 C_c 可以通过下式计算：

$$C_{c\ \mathrm{effective}} = \mathrm{SF} \cdot C_c \quad 其中\ \mathrm{SF} = 1 - \frac{\Delta V_B}{\Delta V_A} \tag{8.3.2}$$

式中，ΔV 为电压开关重叠期间内的电压变化量。根据这一公式，如果 V_A 和 V_B 都是阶跃输入，当二者同相（即 V_A 和 V_B 朝同一个方向开或关）时，SF 等于 0；当二者反相时，SF 等于 2。然而，在纳米领域，实际的压摆率不能再被忽略，信号开关不能再建模为阶跃输入。这样，SF 的取值区间将与 V_A 和 V_B 的压摆率（t_{rA}和 t_{rB}）之比有关。如果接收器的开关阈值为 50% V_{dd}，SF 的取值区间可以大到 [−1，3][38,39]。当 V_A 和 V_B 不同相而且 t_{rB}至少比 t_{rA}小两倍以上时，SF = 3 而不是 SF = 2，出现 V_A 的最差时延情况。这时，总的等效到地电容为 $C_g + 3C_c$，大于阶跃输入时给出的 $C_g + 2C_c$。随着工艺缩小，C_c 通常比 C_g 大（见图 8-2b），因此对 SF 的这一修正对正确估计时序的上下界非常重要。

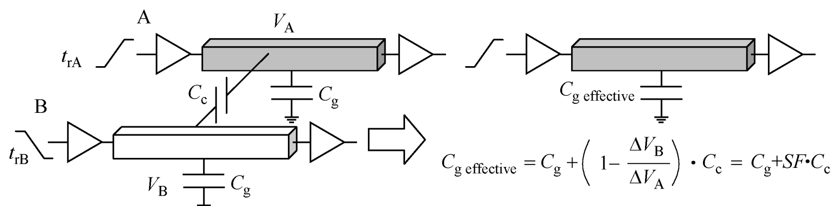


图 8-10 基于开关因子的 *RC* 互连去耦

容性耦合噪声 在基于开关因子的时序分析中，如果互连 B 上不存在开关动作，SF 等于 0。只有当非开关互连的耦合噪声可以被忽略，可以将其作为地节

点对待，这时上述近似才是可行的。然而，在先进的工艺中，由于 C_c/C_g 之比和互连电阻都较大，串扰噪声变得很显著，使得这一假设不再成立。图 8-11a 采用两个耦合 *RC* 互连的集总电路模型来评估生成的噪声（互连建模为 Π 型）。有开关动作并导致噪声产生的互连一般称为攻击线，而受到噪声干扰的互连称为受害线。注意，对容性耦合，仅相邻的互连会影响受害线；次相邻以及更远互连的影响可以忽略。当攻击线出现电压开关动作时，容性噪声只是表现为同一方向。受害线上产生的较大噪声不仅会导致过大的时延不确定性，而且会引入潜在的逻辑误动作。对低噪声容限的设计，例如那些有更高时钟频率、更低供电电压的电路以及采用动态逻辑的电路，逻辑误动作问题显得特别严重。因为高速电路有许多对噪声非常敏感的特性，为了减小昂贵的迭代设计次数并最终确保设计成功，几乎在每个设计阶段都应该考虑串扰噪声的影响。

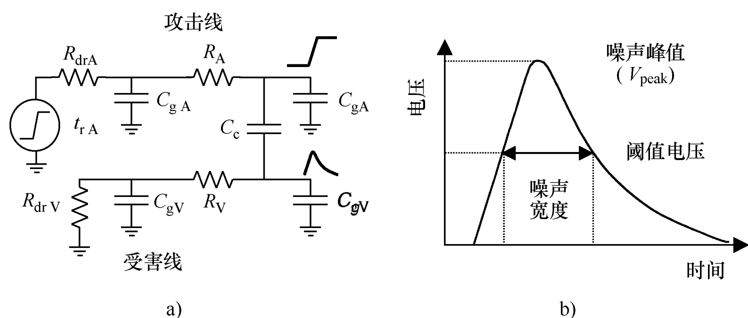


图 8-11　*RC* 分析中的容性耦合噪声

a）一对耦合 *RC* 互连的集总模型　b）主要的噪声表征

如图 8-11b 所示，用于估计噪声影响的度量主要有两个：噪声峰值 V_{peak} 和噪声宽度。V_{peak} 描述了两个网络之间串扰噪声的最大值，其值取决于耦合电容、其他负载电容、寄生电阻、攻击线的开关压摆率、以及受害线驱动器的强度。使用主极点法[40]，V_{peak} 可以近似为

$$\frac{V_{peak}}{V_{dd}}=\frac{t_x}{t_{rA}}(1-e^{-t_{rA}/t_v}) \tag{8.3.3}$$

式中，t_x 和 t_v 分别为攻击线和受害线的趋稳时间，它们的值可以从其他的 *RC* 寄生参数求得解析解[41,42]。类似的解析方案可以参见参考文献［43］和［44］。根据这些理论结果和实际电路的实验结果发现，V_{peak} 对 C_c/C_{gV} 之比的敏感程度明显高于对其他参数的敏感度[42]。实际上，如果受害线是高度电阻性的且攻击线开关非常快，V_{peak} 会逼近电荷分压的上限：

$$\frac{V_{peak}}{V_{dd}}=\frac{C_c}{C_c+C_{gV}} \tag{8.3.4}$$

除了 C_c/C_{gV}，受害线驱动器的电阻 R_{drV} 也在求解 V_{peak} 值中起重要的作用。将这些观察结果并入设计技术中，有助于改善优化效果并抑制不期望的耦合。

噪声振幅峰值 V_{peak} 并不是用于表征噪声的唯一度量。即使 V_{peak} 超过某一阈值，在某种情况下接收器仍然可能不受噪声影响。例如，当噪声宽度非常窄而接收器电容又比较大（即噪声太快不能触发低带宽的接收器）时的情况就是如此。因此，描述大于给定阈值的噪声值持续时间长度，即噪声宽度，通常用来代表噪声的速度。实际设计中采用这一度量的优点就是它可以获得解析解，并且很适用于布线和筛选算法[41]。为了更精确地预测噪声对时序的影响，需要对整个噪声波形加以表征。容性串扰噪声，例如图 8-11b 所示特征描述的容性串扰噪声，可以建模为到达 V_{peak} 值之前是线性上升边，在到达峰值后以指数方式衰减[44]。关于这一模型的进一步讨论详见 8.3.4 节。

8.3.3 *RLC* 互连分析

RC 分析最适合应用于局部层和中间层的电阻性网络，电感效应经常表现在较宽的全局互连中。全局互连用于在功能块之间传输信号、分布时钟参考、为逻辑门供电。为了在高性能电路中能合理地设计这些互连，需要采用表征信号传输特性的 *RLC* 模型和技术。然而，由于以下两个原因，电感效应增加了时序和噪声分析的复杂度。首先，不像容性耦合那样只受最邻近互连的影响，感性耦合的影响会扩展到更远的范围。与电场的行为不一样，磁场在金属表面是非零的，因此互感随着距离衰减得非常慢。其次，因为片上互连结构没有提供意义明确的直流路径以构成电路回路，电路中感性电流返回路径存在不确定性。因此，*RLC* 分析不是局部问题，这意味着必须在较大范围内考虑足够多的相邻互连才能得到正确解。另外，对主要的性能度量，获得简单的解析解也非常困难，因为如式（8.2.3）所示，电感引起的是非单调行为（如振铃和过冲），所以需要匹配更多的矩才能逼近输出特性。

***RLC* 互连时序分析**　与 *RC* 互连相比，*RLC* 互连在电压开关的传输过程中表现完全不同，具有更大的时延、更快的压摆率、有振铃也有过冲。图 8-12a 是 180nm 工艺中电感对典型全局互连上斜坡响应波形的影响。由于感性阻抗的作用，50% V_{dd} 增加，而对时钟边沿和串扰噪声特别重要的信号压摆率则减小了。估计这个效应时，必须考虑两个因素：一方面，一个更陡峭的信号边沿对数字设计更有益，因为状态跳变只需更短的时间；另一方面，信号开关越快，串扰噪声越大（源自容性和感性耦合）。因此，一个最佳的设计应该在给定的噪声约束下实现最小的压摆率。在 *RLC* 中存在有电压振铃和过冲，而在 *RC* 互连中则没有，这是具有感性元件的传输线行为所造成的。这些不期望的特性可能进一步引起不期望的后果：振铃影响时钟信号的稳定，因为较大的振铃会被错误地当作是一次

跳变，从而引起逻辑错误；而电压过冲可能增加功耗并使栅氧甚至整个器件的可靠性退化。除了在 *RLC* 互连的远端存在波形特性外，在沿互连的其他各点上也会出现不期望的行为。例如，图 8-12a 还给出了互连近端的情况，如果驱动器和互连的阻抗不匹配，则电压波形在跳变边沿的中部会出现一个平台。在阈值附近出现这样的电压平台将可能加大驱动器时延，但是这个效应可以通过驱动器尺寸的调整得到优化。

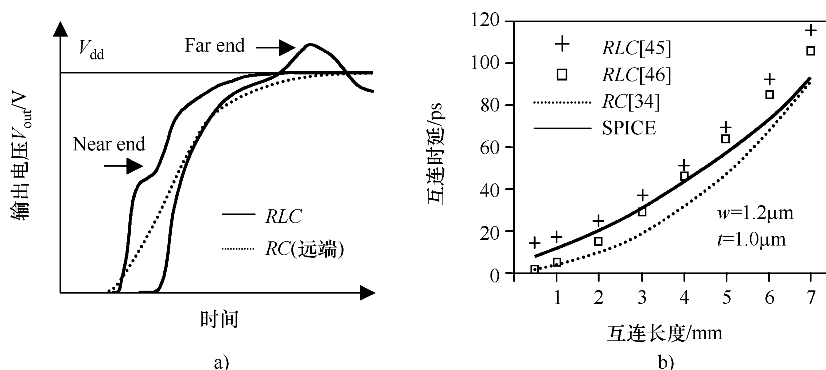


图 8-12　开关行为的比较

a）输出波形比较　b）时延预测比较

与 *RC* 互连分析相似，*RLC* 时序分析可以使用一般的数值技术[27,45]或者解析方法。解析方法的高效率非常适合于全局的 *RLC* 互连布线和优化。为了简化建模时感性耦合的复杂度，首先用等效回路电感 L_{loop}的概念对目标互连与其相邻互连进行去耦。对屏蔽良好的结构，例如时钟线，L_{loop}可以通过计算给出解析解[25,26]。而对于多位的数据总线结构，很容易建立一个可供查阅的查找表，其中 L_{loop}的值是互连配置和输入开关模式的函数[24]。计算获得 L_{loop}的值之后，单个 *RLC* 互连的输出波形可以用矩匹配技术求解[30,46,47]，并且时延度量可以近似为

$$\text{delay} = \frac{e^{-2.9\zeta^{1.35}} + 1.48\zeta}{\omega_n} \tag{8.3.5}$$

式中，ζ 和 ω_n是互连寄生参数的函数[46]。类似的结果还可以参见参考文献［47］和［48］。注意，当 $4L/R^2C>1$ 时（R、L、C 是整个互连长度上总的线寄生参数），会发生信号的振铃和过冲。基于解析时延度量，图 8-12b 比较了由各种 *RC* 和 *RLC* 模型预测的互连时延。总之，考虑到电感影响，信号时延增加了约 15%，并且在互连长度为 2～5mm 范围，*RLC* 模型与 SPICE 结果非常吻合[8]。

除了互连本身的特性外，当邻近的互连都有开关动作时，信号的时延还与多位数据总线的输入向量有关。为了设计一个能查找潜在的信号完整性问题的验证工具，了解攻击线在最坏情况下的开关方向是非常重要的。如果采用简单的 *RC*

模型，业已明确，当受害线开关模式的方向与其相邻互连完全相反时，它将承受最大的有效耦合电容，从而产生最坏的时延情况。考虑长距感性耦合时，对最坏情况估计必须给出全部输入开关模式。如图 8-13 所示，最坏输入向量的两种情况都是对称的。第一种情况（见图 8-13b）对应于所有的邻近邻线同向开关，而且与目标互连方向相反。如果互连以 *RC* 为主，则时延最大。如果更远的互连与目标互连同向开关，因为同相开关产生了最大的回路电感，则电感效应占主导地位（见图 8-13b）对应于导致最坏时延的第二种输入模式。在 180nm 节点，第二种输入模式代表了许多电路实例中的最坏情况[24]。但是，实际中最坏情况的开关模式不仅与 *RLC* 参数有关，还取决于工艺。

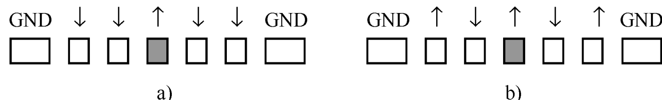


图 8-13 最坏情况可能的输入向量（箭头↑表示向上开关；箭头↓表示向下开关）
a）模式 1：容性耦合为主 b）模式 2：感性耦合为主

总之，只要输入向量合适，*RLC* 模型会给出信号时延估计的一个上限，而 *RC* 模型一般会给出压摆率计算的一个下限。

感性耦合噪声 工作频率和全局互连长度的持续增加不仅会导致显著的电感效应；而且会加重纳米领域的串扰噪声。在容性串扰和感性串扰之间，存在两点基本差异：

（1）噪声的极性。在容性耦合中，串扰噪声（$C(\mathrm{d}V/\mathrm{d}t)$）的方向总是与“攻击线”开关的方向相同。然而，感性耦合通过返回电流产生的噪声（$L(\mathrm{d}I/\mathrm{d}t)$），则与攻击线开关的方向相反，并且发生的比 $C(\mathrm{d}V/\mathrm{d}t)$ 更及时。因此，对于给定的攻击线开关动作，感性噪声的极性一般与容性耦合的极性相反并且出现的时间更早，如图 8-14a 所示。对第一阶和第二阶邻线，正的（即容性耦合）和负的（即感性耦合）噪声峰值都与 *RLC* 预测结果比较吻合。然而，这些相反的因素抑制着相邻互连间耦合噪声的幅度。

（2）耦合范围。因为感性耦合产生的返回电流延伸的距离和范围很广，即使更高阶的受害线也可能受到 *RLC* 串扰噪声的影响，如图 8-14a 所示。虽然容性耦合随着距离的增加快速衰减，而对于非相邻互连之间的感性耦合却不能忽略。实际上，由于第二阶互连上没有相反的容性耦合噪声，使得第二阶互连的最大感性噪声（负的峰值）会大于第一阶互连的感性噪声。

由于这些耦合机制的竞争特性，没有电路仿真器的辅助很难精确预测总的 *RLC* 串扰噪声行为。图 8-14b 显示了峰值噪声和互连长度之间的复杂关系。较宽互连的 V_{peak} 值更趋向于感性效应，并且表现出非单调的相关性；在 *RC* 占主导的

较窄互连中，V_{peak}值随着长度的增加而单调增加。另外，特别是在较长的平行数据总线结构下，感性耦合噪声比容性耦合噪声更严重且更难控制。实际应用中，为了限制返回电流，在每两根或每四根信号线之间插入回路电源线和地线。然而，即使采用这种预防性的方式，感性噪声仍然会跨过屏蔽区对受害线产生影响[49]。因此，应该在设计的早期阶段就采用版图（比如参考文献[50]中的内容）或电路设计技术预防电感效应的影响，而不是依靠在后期验证阶段的昂贵分析工具。

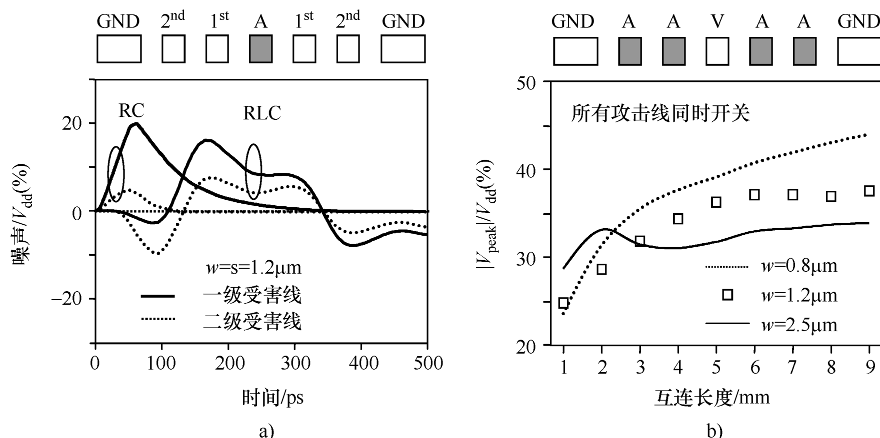


图 8-14　180nm 铜互连工艺（互连长度 = 3mm；A 代表“攻击线”；V 代表受害线的串扰噪声比较

a）噪声波形比较　b）不同宽度的峰值比较

8.3.4　考虑噪声耦合效应的时序分析

如果设计中采用的金属节距更近、纵横比更大和工作频率更高，就更容易受到互连耦合效应的影响。过大的串扰噪声会引起受害线网络出现错误的开关动作；即使是较小的噪声也能极大地改变受害线的时延，引起时延抖动。当某一级（包括门和互连）的时序，由于附近门的开关活动形成时延抖动，就会导致这种时序的不确定。由于 CMOS 逻辑具有恢复特性，只有大于接收器开关阈值的噪声毛刺才会引起逻辑功能故障。相比之下，对于较短的互连（小于 500μm），时延抖动则更加普遍，并且很容易大于标称时延的 20 ~ 30%[51]。图 8-15a 所示为各工艺级别中 3mm 全局互连时延不确定性的增加情况。该例中，最坏情况是，纳米领域归一化的时延抖动已经接近 80%（注意，在从低到高的跳变中，时延抖动更严重，因为此时受害的 PMOS 比攻击的 NMOS 更脆弱）[51]。像这样巨大的时延不确定性给具有更严格时序预算的高性能设计提出了严酷的挑战。

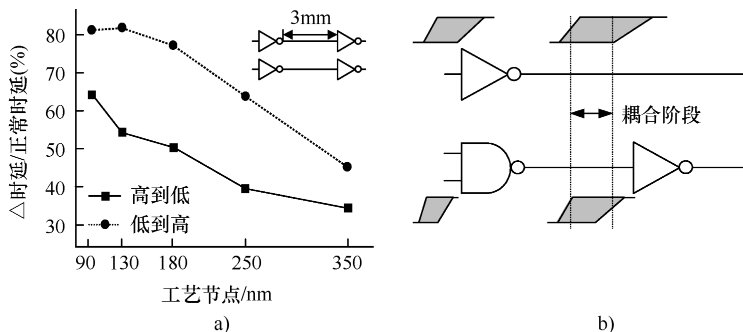


图 8-15 使用开关窗口法进行的考虑噪声耦合效应的时序分析

a）噪声引起时延抖动的发展趋势 b）基于开关窗口的时序分析

（阴影区代表可能开关活动的时间窗口）

为了避免最坏耦合情况下的芯片时序故障，最重要的是在进行静态时序分析（Static Timing Analysis，STA）时应考虑时延抖动，并预留足够的容限以能够承受时延的波动。在 *RC* 分析中，广泛采用的一种计算最早和最迟串扰时延的方法是，利用开关因子（其取值区间为［0，2］，或更精确的是［－1，3］），将各个关键路径上的耦合电容折算为等效到地电容。这一技术既稳妥又容易实现，而且不需要邻近网络的信息。然而，虽然这种简化降低了分析的复杂度，但是它会导致过分悲观的估计或过分严格的布线间距，因为事实上串扰噪声只是当攻击线和受害线同时开关时才会影响信号时延。如果攻击线和受害线的开关没有重叠，那么就不需要考虑时延抖动。因此，常规的方法往往过高地估计了时序的界限并浪费了计算时间。

改善噪声敏感 STA 精度的关键是应该包括关于信号网络的时域和功能信息。这可以通过引入开关窗口（又称时序窗口）的概念加以实现。这一窗口是指节点发生跳变的那段时间，如图 8-15b[52,53] 所示。两个耦合节点的信号时延只有当它们开关窗口重叠导致串扰噪声时才可能出现变动；否则信号时延不受时延抖动的影响（见图 8-15b）。剩下的问题就是如何确定每个节点的开关窗口。为此，虽然可以直接采用攻击线的时序信息，但是攻击线的开关窗口又要受到受害线开关窗口的影响，这就是典型“鸡和蛋”的问题。虽然有多种方法可以化解这个循环，一个比较普遍的方案就是反复的计算迭代[53,54]。首先，我们可以为网络假定一个初始耦合情况（例如最坏情况耦合），运行时延引擎以估计每个节点的时延界限（即开关窗口），再根据开关窗口之间的相互关系重新估计耦合情况。重复这个过程直到时序窗口收敛[52]。

使用基于开关窗口的噪声敏感时序分析可以极大地减小时延界限估计的悲观性。在这一框架下，根据对串扰和时延抖动基本关系的认识，在改善时序窗口计

算的效率和精度方面已经取得不少进展。如图 8-16a 所示，受害线时延在串扰噪声出现时会发生变化，因为噪声波形耦合到受害线后使得初始的电压传播发生失真。随着噪声注入位置的不同，可以观察到时延的不同变化。如果知道开关电压和噪声毛刺的标称特性，又知道攻击线和受害线的输入时序信息，就可以采用波形重叠方法预测这一级输出的时延抖动。时延变化曲线（Delay Change Curve，DCC）验证了这一想法。DCC 表示的是时延与攻击线和受害线之间信号相对到达时间的关系[51,55,56]。图 8-16b 所示为 0.35μm 工艺中长度为 6mm 的全局互连的 DCC 测量结果[51]。通过使用 DCC，与传统的使用开关因子估计的时延区间相比，输出时序窗口已经准确地变小，结果与 DCC 中的峰峰值比较吻合。实际中，可以从分析波形重叠中有效地获得 DCC[51]。

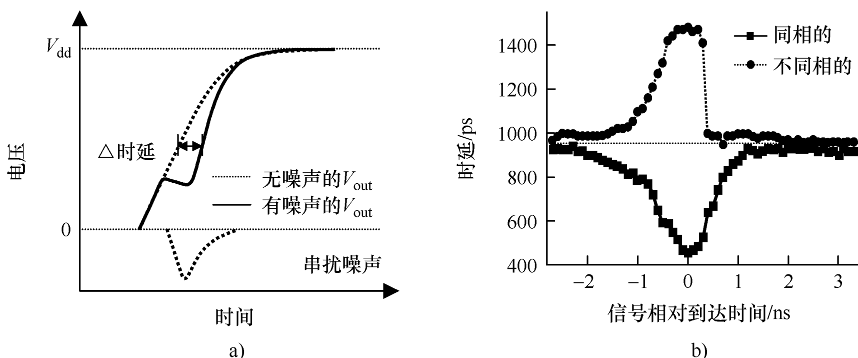


图 8-16　用于时序窗口估计的 DCC

a）波形重叠引起的时延变化　b）测量获得的时延变化曲线

8.4　信号完整性设计技术

时序很关键的互连，如时钟和全局信号总线，一般会针对时延、上升边和噪声进行优化设计。优化过程包括在不同设计层次考虑信号完整性问题的预防、分析和修复。在物理设计的前期阶段，在可能的情况下会增加额外的布线限制以避免过量的噪声。像插入转发器这样的电路设计技术在改善信号质量和互连性能方面是有效的。在完成布线以后，通常需要最终调整互连布线以解决信号完整性问题。每种这样的技术都有其优缺点。通过增加额外的布线限制以预防信号完整性问题的方法很容易实现，但是由于芯片尺寸和功耗等问题，成本会变得很高。由于在这一设计阶段估计的串扰噪声较粗略，不仅不能修正所有的错误，并且还可能会带来一些不必要的改动。在布线后再对互连进行调整可以得到更精确的结果，但对整个设计流程而言有点太晚而且可能会出现不收

敛的问题。总之，如果既要将信号完整性维持在可控的范围内，同时又对成本和成品率的影响最小，就必须在不同的设计阶段使用所有上述方法。本节我们将讨论版图设计前和版图设计后优化这种物理互连结构以及针对信号完整性敏感的设计策略。

8.4.1 物理设计技术

时延和上升边与物理互连结构（即长度、宽度、间距、驱动器尺寸等）的关系非常密切。耦合噪声的幅度强烈依赖于互连布局的情况、彼此之间的远近程度以及邻近线的跳变活动，而跳变活动又与驱动强度和负载电容相关。设计者可以使用不同技术优化其中每个特性，例如，采用噪声约束布线、网络的重排序、门尺寸的缩放、互连结构的优化等。

噪声约束布线 串扰大小明显取决于布线。近几年，CAD 工具在布线阶段已经包括了信号完整性的预防和校正技术，称为噪声约束或噪声免疫/回避布线。这是一个 NP 难题并且不能精确求解，因此需要采用启发式试探的方式。首先，基于常规的布线方案构建一个初始解，然后估计每个网络中的串扰。下面是一个串扰噪声估计的实例：用一个预先定义的边界（如预先布好的电源/地网格）将设计分为几个不同的区域，且假定不同区域之间的耦合为零。对容性耦合，假设版图设计只控制耦合电容（间距），因为其他参数（驱动器强度、负载电容、输入波形等）要么不能修改，要么就是修改不好。假定容性耦合噪声与耦合电容成正比。对区域内的每个网络，计算网络对电容耦合噪声的敏感度，采用容性串扰系数加以衡量：

$$C_i = \sum_{j \neq i} C_{ij} \tag{8.4.1}$$

式中，C_{ij}为网络 i 和 j 之间的耦合电容。耦合电容在最近互连之外迅速减小，因此在上面的求和中一般只需包括第一阶邻线。上面的度量忽略了这样一种情况，即如果两个网络在不同时间开关，它们的串扰噪声就不会影响电路性能。但是，表征所有可能的开关情况需要耗费大量的时序分析，而时序分析又转而取决于串扰。因此，在最坏情况时，可以使用邻线所有耦合电容的总和来代表总的串扰噪声。一个变通方案就是简单地将串扰估计中的初步时序信息相加，以尽量避免大量的过份估计。噪声约束可以设定为

$$C_i < \overline{C_{\max}} \tag{8.4.2}$$

如果发生违例情况，可采用增加间距、插入屏蔽和网络重排序等补偿技术改善设计。如果采取这些措施之后某个区域还有违例的情况，这个区域的一些网络可以被移到其他区域重新布线。因为这是粗略的串扰噪声估计，噪声约束布线很容易造成过分的或欠缺的设计。

如 8.3 节讨论的，感性耦合随着时钟频率的升高和工艺的缩小变得更重要。与上述模型相类似，可以用感性串扰系数 K_i 估计网络对感性噪声的敏感度[57]：

$$K_i = \sum_{j \neq i} lk_{i,j} \quad \text{并且}\ K_i \leqslant \overline{K_{\max}} \tag{8.4.3}$$

式中，$k_{i,j}$ 是网络 i 和 j 之间的感性耦合系数；l 是区域长度。感性噪声的耦合距离远大于容性耦合，因此相比容性耦合，式（8.4.3）中需要包括更多的网络。另外，需要有像电源/地网格那样现有的边界给问题提供约束。然而，对感性耦合，这种电压/地屏蔽规则有时会低估实际的感性噪声。

减小邻线开关活动的另一项技术是故意让驱动器错位。这个技术使总线内的互连有正常的和偏移的两种时序。这样就不会出现相邻互连会同时开关的情况，不管正常的或被延迟的信号都将受到很小的邻线耦合影响。互连时序的偏移可以用反相器链或两相时钟方案实现。虽然这种技术在有时间偏移的互连上引入了时延，但是因为占主导位置的串扰时延被抑制，对于有偏移互连的总线，总时延还是减小了。

驱动器尺寸缩放　现在从受害线和攻击线驱动器的角度分析驱动器尺寸对信号完整性的影响。从直觉上讲，如果增大受害线驱动器尺寸，将会增加其有效电导，这将使网络上的信号更稳定。另一方面，如果增大攻击线驱动器的尺寸，它在受害线线上造成噪声的量也会增加。因此，增加驱动器尺寸对串扰噪声有双重影响。大尺寸驱动器互连上出现的噪声减小了，但是它在邻近互连上造成的噪声将增加。对于给定的设计，图 8-17 中定量描述了驱动器尺寸和其他互连参数（C_c、C_{a1} 等）是如何分别影响串扰噪声[58]的。图 8-17a 所示为耦合噪声模型，其中攻击线和受害线被分为三个区域：耦合处之前的互连片段、耦合处和耦合处之后的片段。实际电路中每个模型参数的噪声敏感度如图 8-17b 所示。

互连调节　互连调节应该与晶体管尺寸的缩放同时进行。对 RC 互连，通过调节减小互连时延最有效的方式是增加互连宽度。当宽度增加时，更宽互连电阻的减小比总电容的增加更快（当耦合电容占主导地位时），所以一般时延会更小。因为耦合串扰的两个主要原因是耦合电容和邻近开关情况，减小噪声最有效的方法是增加线间距并加大总线的阶数。针对 RC 网络调节的简单经验法则是

（1）为了减小时延，增加互连宽度（比增加线间距更有效）或插入转发器。

（2）为了减小串扰，增加线间距（比增加互连宽度更有效），网络重排序，或插入转发器。

对于感性网络，互连调节变得更加棘手。加宽互连会在主导互连上导致更大的电感，引入感性振铃和额外的时延。另外，互连间距增加引起的回路电感的增

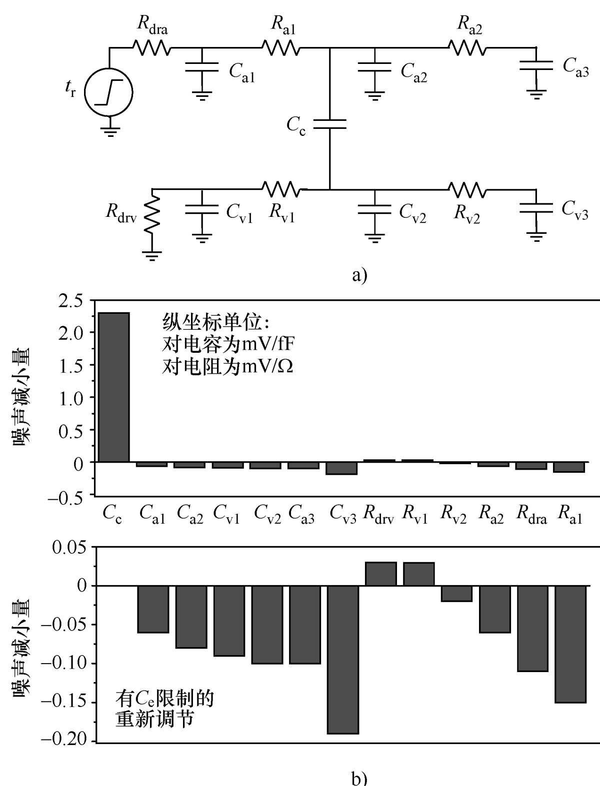


图 8-17 a）耦合噪声模型 b）模型参数对峰值噪声的敏感度（来源于参考文献［58］）

加将抵消一部分耦合电容减小的收益。当需要较宽的互连驱动大负载时，可能需要将它们分成用 VDD/GND 屏蔽隔开的多个小“指条”。图 8-18 所示为一个全局时钟结构的例子，其中互连被分成三条互连并被完全屏蔽。由图可见，对相同的布线面积，通过优化互连结构可以获得很大的性能改善（这里用时延和上升边表示）。互连的信号返回比定义为总的时钟互连宽度（$T_{CLK}=N\cdot W_{CLK}$）与总的地屏蔽互连宽度（$T_{GND}=(N+1)W_{GND}$）之比，在时钟频率为 2GHz、噪声限制为≤10% 时，对于全屏蔽时钟结构，应遵循下列规则[26]

$$\text{最佳时延}: T_{CLk}:T_{GND}\approx 0.9\sim 1 \qquad S:W_{GND}\approx 0.4\sim 0.5 \tag{8.4.4}$$

$$\text{最佳功耗}: T_{CLk}:T_{GND}\approx 0.8 \qquad S:W_{GND}\approx 0.7 \tag{8.4.5}$$

这些比率将随着频率的升高和互连分支数的增加而减小（意味着 W_{GND} 增加），这是因为地线返回电阻的重要性增大了。设计 *RLC* 互连的简单经验法则是：

（1）至少提供与信号尽量接近的返回路径（$T_{GND}\geqslant T_{CLK}$）。

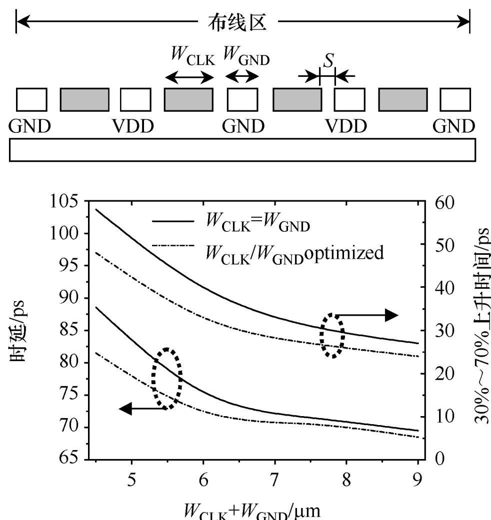


图 8-18　通过互连结构的优化实现性能的改善

（2）要采用比最小间距大的间距，因为由此引起耦合电容的减小量大于回路电感的增加量。

在高速总线内使用 VDD/GND 作为屏蔽互连是限制信号线耦合的最通用设计技术，但这是以布线面积的增加为代价。它有效地减小了容性耦合以及相关的时延不确定性。对于 *RLC* 网络，地线屏蔽提供了更近的电流返回路径并减小了回路电感。由于在这一对互连附近产生的磁场方向相反而相互抵消，因此也就减小了感性噪声的产生。图 8-19 给出了屏蔽密度对信号噪声和时延的影响。如图 8-19所示，电感占主导地位互连中（$W=2.5\mu m$，$S=1.25\mu m$），噪声与信号互连数之间成线性关系，因为在控制噪声效果方面，屏蔽技术对感性耦合的控制不如对容性耦合有效。从时延优化的角度，对 $W=0.8\mu m$ 互连，在每三条互连之间放置一个屏蔽线以后，时延基本保持不变。对感性更强的 $W=2.5\mu m$ 互连，随着屏蔽线之间互连数目的增加，时延持续变差。一般情况下，在每两条到四条信号线之间放置屏蔽线会达到最好的面积利用率。对未来更高工作频率下的工艺，可能需要专用的接地等位面以减小感性耦合。这一技术经常被 PCB 和封装设计者使用，但是对当前工艺的集成电路芯片设计者而言，它太昂贵了。在实际设计时，诸如互连拥塞、噪声和电源线上的 *IR* 压降等，也是确定互连几何尺寸时需要考虑的重要因素。

在一个高性能微处理器的设计实例中[58]，对所有的受害线网络应用了各种噪声回避技术，给出了在 48000 条较长互互连上最终噪声减少的平均百分比。从这个实例中可以看出，最有效的减小噪声技术是互连间距，但是其成本也高。另

外，受害线驱动器尺寸的缩放也有相当的效果，而互连尺寸的缩放已经被证明效果最差。当然，具体一项噪声减小技术的有效性还要取决于该网络特定的互连/驱动器特性。

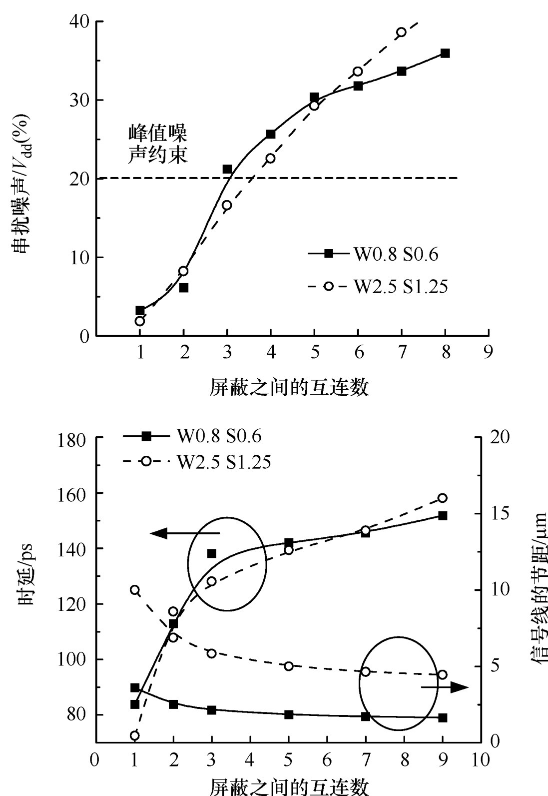


图 8-19　最佳屏蔽方案是在每 2 ~4 条信号线之间放置一条屏蔽线

8.4.2　电路技术

插入转发器　插入转发器（缓冲器）技术对减小长互连的长时延是很有效的，但是要付出的代价是芯片面积和功耗的增加。该技术是将长互连分割，并在各个片段之间插入驱动器（转发器）（见图 8-20），可以使互连长度与时延关系从平方降为线性，从而极大地缓解了长互连的时延问题。因为 CMOS 驱动器的特性，插入转发器也极大地改善了远端驱动器的信号压摆率。因为实际情况下，驱动器和接收器的尺寸与转发器的尺寸不同，除了路径上第一个和最后一个片段外，转发器一般以均匀间隔插入。另外，在实际设计中，转发器一般也是采用级联反相器对实现，以尽可能地减小时延。

对 *RC* 互连，普遍采用的最优缓冲器尺寸表达式是 Bakoglu 方程[34]。最优转

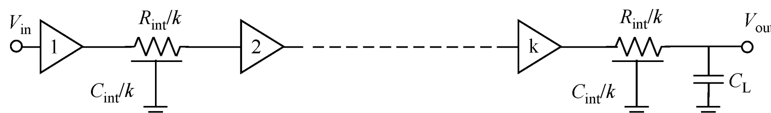


图 8-20　通过插入转发器减小 RC 互连时延

发器的个数是

$$k_{opt} = \sqrt{\frac{0.4R_{int}C_{int}}{0.7C_0R_0}} \tag{8.4.6}$$

式中，R_0 和 C_0 分别是最小尺寸转发器的输出电阻和输入电容。转发器的尺寸为

$$h_{opt} = \sqrt{\frac{R_0C_{int}}{R_{int}C_0}} \tag{8.4.7}$$

然而，由式（8.4.7）计算得到的结果经常太大以致于不太现实。典型的标准单元库中包括的是最小尺寸 50～100 倍的反相器或缓冲器，而式（8.4.7）给出的结果范围则高达最小尺寸的 400～700 倍。实际上，为了承受较大的时延，一般是插入适当的转发器而不是最优的转发器。参考文献［59］中给出的表达式优化的是加权时延-面积积，而不是纯粹的时延度量。其结果比式（8.4.7）给出的小了 50%～60% 量级：

$$W_{optarea} = \frac{0.541}{R_{int}C_{int}} - 0.231R_DC_{in} - 0.126R_{int}C_{int} + \sqrt{0.053R_D^2C_{in}^2 + 0.058R_DC_{int}R_{int}^2C_{int}^2 + 1.708R_DR_{int}C_{int}} \tag{8.4.8}$$

基于式（8.4.8）的时延较高，但是面积和功耗的成本相对较小，如果再优化能量-时延积，其值将会更小。

如果将一条互连分成更多的较短片段，则网络具有更强的抗噪声能力。转发器的插入减小了互连的平行长度，从而强烈地影响着串扰噪声。图 8-21 表明了采用和未采用转发器两种情况下噪声对受害线网络的影响。上面的互连是攻击线网络，下面的互连是受害线。如图所示，插入缓冲器后在插入缓冲器的输入端出现的噪声脉冲（见图 8-21b）比图 8-21a 中接收器输入端的噪声脉冲要小。由于缓冲器的特性，这个较小的噪声很容易得到抑制。对感性噪声耦合，由于原先的电流返回路径现在经由转发器返回，这就缩短了电流返回路径长度，使得电流回路变小，从而感性耦合也变小。不过，因为 LC 互连的时间常数 $\sqrt{LC}$ 与互连长度近似成线性关系，而不是 RC 情况下的平方关系，插入转发器对 RLC 互连时延的减小基本没有作用。

在相邻互连上分别插入转发器时，使插入的转发器相互错开可以使耦合电容对时延和串扰噪声的影响最小（见图 8-22）[8]。转发器的偏移错开放置

要求是使每个门位于它邻近门互连负载的中间。因为潜在的最坏邻近互连同时开关情况针对的只是受害线长度的一半，所以有效的开关因子被限制到 1 以内。而由于对称的缘故，受害线的另一半将经历属于最好情况的邻线开关活动。采用错开的转发器，可以极大地减小由邻近互连开关活动引起的时延不确定性。

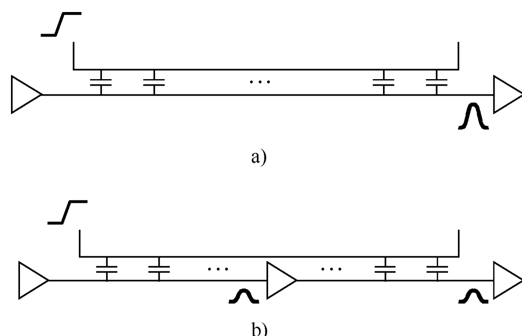


图 8-21 转发器有助于抑制受害线网络上的耦合噪声
a）无转发器情况 b）有转发器情况（来源于参考文献［63］）

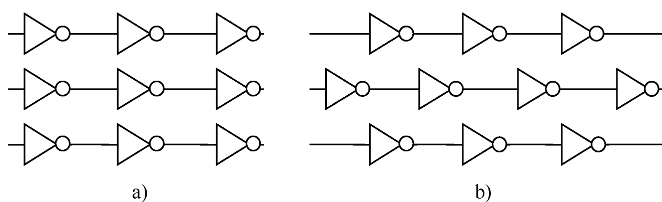


图 8-22 错开转发器以减小开关模式引起的时延波动

最新研究表明，转发器占用了越来越大的面积、功耗和设计资源，并且关键在于它们对性能的改善是有限度的。最近几年出现了一些关于更有效地驱动长互连方面的创新设计，包括可恢复的提升器[60,61]。与转发器不同，提升器是沿着互连加上去以局部地增强所传输的信号，它并不改变互连布线。提升器能够即时感受到互连上电压的跳变，提供一个额外的电流提升以加速跳变。提升器的电路原理图和时序图如图 8-23 所示。该提升电路有两个错位的反相器，用以在跳变尚未到达正常反相器之前检测到跳变。由它们再驱动一条反馈路径从而局部加速开关信号。另外，又引入一个 Muller-C 电容元件以防止在 VDD 和 GND 之间形成直接通路。这个提升电路的一个优点是它的性能对布局的变化不敏感，并且几乎可以被放置在总线上的任何一点；这样基本的信号布线约束对它们的影响很小。另外，它不改变信号的极性，支持双向传输和多驱动源

配置。一些实验已经表明，使用恢复提升器技术能驱动更长的互连，并且比转发器方案节省面积和功耗。然而，设计提升器的主要难点是潜在的亚稳定性问题，这是具有正反馈的电路所固有的，其禁止具有不同脉冲宽度的信号传输通过。

图 8-23　提升电路实例

保持器电路　在微处理器的关键性能单元和其他高性能 VLSI 电路中经常采用动态门。与静态 CMOS 门不同，在动态节点上由于噪声引起的电荷损失是不能恢复的，从而动态门比静态 CMOS 门更易受到噪声的伤害。采用称作保持器的上拉和/或下拉器件形成静态通路，可以避免动态门节点的浮空（见图 8-24）[62]。保持器电路可以恢复由耦合噪声、电荷共享和亚阈值漏电流引起的电荷损失。然而，随着噪声和漏电流的不断增加，必须相应地加大保持器的尺寸，这会使动态电路的性能明显下降。

图 8-24　保持器可以恢复动态电路损失的电荷

差分信号　与单端信号相比，差分信号自然就具有更鲁棒的抗噪声能力。差分信号的基本概念如图 8-25 所示，其中两条紧耦合的互连以差分的方式传送

图 8-25　差分信令

数据。在接收处，将这两个信号加以比较以确定它们的逻辑极性。差分信号可以用电压模式和电流模式实现。

差分信号机制具有高抑制共模干扰能力，这些干扰包括串扰噪声和电源轨波动。它也提供了一些单端信号所不具备的其他优点，包括：

（1）对每条信号线，它有一条很邻近的内建返回路径，因此几乎没有耦合到其他网络的噪声；

（2）因为具有的抗噪声能力高，就可以采用较低的信号摆幅，达到减小功耗的目的——已经证明，摆幅低至 200mV 的情况也能正常工作；

（3）信号与电源轨及其相关噪声隔离，使差分接收器受到的所有电源噪声都是共模形式，而差分接收器又都设计有非常好的共模抑制能力。因此，差分信号可以工作在噪声更大的环境中，并且可以工作的信号速率比单端电路更高。然而，实现这个技术需要付出巨大的代价。因为 N 个信号需要 $2 \times N$ 条布线通道。发送器和接收器也要有大量的设计管理，并且容易受到时钟错位和抖动变化的危害。

将差分信号与电流模逻辑结合的技术已经广泛应用到片外互连中。图 8-26 显示的是工作于 6.4GHz 的电流模双向差分信号实例。下一代的高性能电路，当芯片工作频率升高或电源噪声产生的错位太大，导致感性噪声是片上的主要问题时，这一技术会成为片上互连一个很有前途的解决途径。

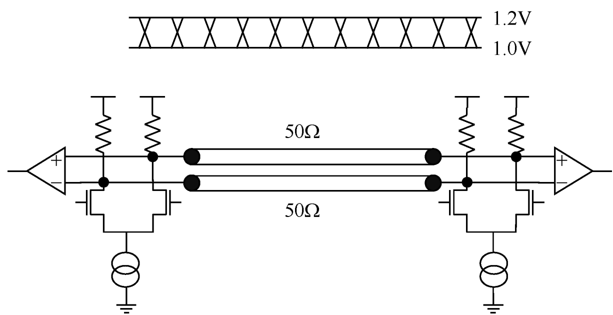


图 8-26　双向差分信号实例

8.5　总结

信号完整性问题，包括串扰噪声、信号过冲和电源噪声，已经成为当前设计中的重要问题。为了保证足够高的芯片成品率并且避免过高的设计成本，应该在设计流程中，尽可能早地考虑信号完整性问题，而不要再只是依赖于版图设计

后昂贵的全芯片验证和纠正技术。为此，通常的时序驱动方法学已经进化到将信号完整性约束加进每个设计阶段中。这些约束可以采用解析模型表征，从物理上将设计空间和电路性能相关联。本章给出的各种电路和物理设计技术为信号质量的改善提供了多种进一步的选择。对从事实际设计工作的设计者，最重要的是，不断保持对可能的信号完整性问题的高度重视，构建一个抑制潜在故障的防火墙。

参 考 文 献

[1] International Technology Roadmap for Semiconductors, *http://public.itrs.net.*

[2] Berkeley Predictive Technology Models, *http://www-device.eecs.berkeley.edu/~ptm.*

[3] D. A. B. Miller and H. M. Ozaktas, Limit to the bit-rate capacity of electrical interconnects from the aspect ratio of the system architecture, *J. Parallel Distribut. Comput.*, Vol. 41, No. 1, pp. 42–52, Feb. 1997.

[4] A. Deutsch et al., Bandwidth prediction for high-performance interconnections, *IEEE 50th Electronic Components and Technology Conference*, pp. 256–266, 2000.

[5] A. Deutsch et al., When are transmission-line effects important for on-chip interconnections, *IEEE Trans. Microwave Theory Tech.*, Vol. 45, No. 10, pp. 1836–1846, Oct. 1997.

[6] Y. I. Ismail, E. G. Friedman, and J. L. Neves, Figures of merit to characterize the importance of on-chip inductance, *IEEE Trans. VLSI Syst.*, Vol. 7, No. 4, pp. 442–449, Dec. 1999.

[7] P. J. Restle, A. E. Ruehli, S. G. Walker, and G. Papadopoulos, Full-wave PEEC time-domain method for the modeling of on-chip interconnects, *IEEE Trans. Comput. Aided Des. Integrated Circuits Syst.*, Vol. 20, No. 7, pp. 877–887, July 2001.

[8] Y. Cao et al., Effects of global interconnect optimizations on performance estimation of deep submicron designs, *Proceedings of the International Conference on Computer Aided Design*, pp. 56–61, Nov. 2000.

[9] T. Sakurai and K. Tamaru, Simple formulas for two- and three-dimensional capacitances, *IEEE Trans. Electron Devices*, Vol. 30, pp. 183–185, 1983.

[10] J.-H. Chern, J. Huang, L. Arledge, P. -C. Li, and P. Yang, Multilevel metal capacitance models for CAD design synthesis systems, *IEEE Electron Device Lett.*, Vol. 13, pp. 32–34, 1992.

[11] S.-C. Wong, G.-Y. Lee, and D.-J. Ma, Modeling of interconnect capacitance, delay, and crosstalk in VLSI, *IEEE Trans. Semicond. Manuf.*, Vol. 13, No.1, pp. 108–111, Feb. 2000.

[12] W. Jin, Y. Eo, W. R. Eisenstadt, and J. Shim, Fast and accurate quasi-three-dimensional capacitance determination of multilayer VLSI interconnects, *IEEE Trans. VLSI Syst.*, Vol. 9, No. 3, pp. 450–460, June 2001.

[13] E. You et al., Parasitic extraction for multimillion-transistor integrated circuits: methodology and design experience, *IEEE Custom Integrated Circuits Conference*, pp. 491–494, 2000.

[14] D. Sylvester, J. C. Chen, and C. Hu, Investigation of interconnect capacitance characterization using charge-based capacitance measurement (CBCM) technique and three-dimensional simulation, *IEEE J. Solid-State Circuits*, Vol. 33, No. 3, pp. 449–453, Mar. 1998.

[15] A. Deutsch, Electrical characteristics of interconnections for high-performance systems, *Proc. IEEE*, Vol. 86, No. 2, pp. 315–355, Feb. 1998.

[16] Y. Eo, W. R. Eisenstadt, and J. Shim, *S*-parameter-measurement-based high-speed signal transient characterization of VLSI interconnects on SiO_2–Si substrate, *IEEE Trans. Adv. Packag.*, Vol. 23, No. 3, pp. 470–479, Aug. 2000.

[17] A. E. Ruehli, Inductance calculations in a complex integrated circuit environment, *IBM J. Res. Dev.*, pp. 470–481, Sept. 1972.

[18] E. B. Rosa and F. W. Grover, *Formulas and Tables for the Calculation of Mutual and Self-Inductance*, U.S. Government Printing Office, Washington, DC, 1916.

[19] X. Qi et al., On-chip inductance modeling and *RLC* extraction of VLSI interconnects for circuit simulation, *Proceedings of Custom Integrated Circuits Design Conference*, pp. 487–490, 2000.

[20] K. Gala et al., On-chip inductance modeling and analysis, *Proceedings of Design Automation Conference*, pp. 63–68, 2000.

[21] Z. He, M. Celik, and L. Pileggi, SPIE: sparse partial inductance extraction, *IEEE Design Automation Conference*, pp. 137–140, 1997.

[22] A. Devgan, J. Hao, and W. Dai, How to efficiently capture on-chip inductance effects: introducing a new circuit element *K*, *IEEE International Conference on Computer Aided Design*, pp. 150–155, Nov. 2000.

[23] M. W. Beattie and L. T. Pileggi, On-chip induction modeling: basics and advanced methods, *IEEE Trans. VLSI Syst.*, Vol. 10, No. 6, pp. 712–729, Dec. 2002.

[24] Y. Cao et al., Effective on-chip inductance modeling for multiple signal lines and application on repeater insertion, *IEEE Trans. VLSI Syst.*, Vol. 10, No. 6, pp. 799–805, Dec. 2002.

[25] B. Krauter and S. Mehrotra, Layout based frequency dependent inductance and resistance extraction for on-chip interconnect timing analysis, *IEEE Design Automation Conference*, pp. 303–308, 1998.

[26] X. Huang, P. Restle, T. Bucelot, Y. Cao, and T. -J. King, Loop-based interconnect modeling and optimization approach for multi-GHz clock network design, *IEEE J. Solid-State Circuits*, Vol. 38, No. 3, p. 457–463, Mar., 2003.

[27] C.-K. Cheng, J. Lillis, S. Lin, and N. Chang, *Interconnect Analysis and Synthesis*, Wiley, New York, 2000.

[28] Y. Cao, X. Huang, D. Sylvester, T. King, and C. Hu, Impact of frequency-dependent interconnect impedance on digital and RF design, *IEEE International ASIC/SoC Conference*, pp. 438–442, Sept. 2002.

[29] F. Dartu, N. Menezes, and L. T. Pilleggi, Performance computation for precharacterized CMOS gates with *RC* loads, *IEEE Trans. Comput. Aided Des. Integrated Circuits Syst.*, Vol. 15, No. 5, pp. 544–553, May 1996.

[30] X. Huang, Y. Cao, D. Sylvester, T. King, and C. Hu, Analytical performance models for *RLC* interconnects and applications to clock optimization, *IEEE International ASIC/SoC Conference*, pp. 353–357, Sept. 2002.

[31] L. T. Pilleggi and R. A. Rohrer, Asymptotic waveform evaluation for timing analysis, *IEEE Trans. Comput. Aided Des.*, Vol. 9, No. 4, pp. 352–366, Apr. 1990.

[32] J. Qian, S. Pullela, and L. Pilleggi, Modeling the "effective capacitance" for the *RC* interconnect of CMOS gates, *IEEE Trans. Comput. Aided Des. Integrated Circuits Syst.*, Vol. 13, No. 12, pp. 1526–1535, Dec. 1994.

[33] A. B. Kahng and S. Muddu, New efficient algorithms for computing effective capacitance, *International Symposium on Physical Design*, pp. 147–151, 1998.

[34] H. B. Bakoglu, *Circuit, Interconnections, and Packaging for VLSI*, Addison-Wesley, Reading, MA, 1990.

[35] W. C. Elmore, The transient analysis of damped linear networks with particular regard to wideband amplifiers, *J. Appl. Phys.*, Vol. 19, No. 1, pp. 55–63, 1948.

[36] R. Gupta, B. Tutuianu, and L. T. Pileggi, The Elmore delay as a bound for *RC* trees with generalized input signals, *IEEE Trans. Comput. Aided Des. Integrated Circuits Syst.*, Vol. 16, No. 1, pp. 95–104, Jan. 1997.

[37] T. Sakurai, Closed-form expressions for interconnection delay, coupling, and crosstalk in VLSIs, *IEEE Trans. Electron Devices*, Vol. 40, No. 1, pp. 118–124, Jan. 1993.

[38] P. Chen, D. A. Kirkpatrick, and K. Keutzer, Miller factor for gate-level coupling delay calculation, *Proceedings of the International Conference on Computer Aided Design*, pp. 68–74, Nov. 2000.

[39] A. B. Kahng, S. Muddu, and E. Sarto, On switch factor based analysis of coupled *RC* interconnects, *IEEE Design Automation Conference*, pp. 79–84, 2000.

[40] M. Kuhlmann and S. S. Sapatnekar, Exact and efficient crosstalk estimation, *IEEE Trans. Comput. Aided Des. Integrated Circuits Syst.*, Vol. 20, No. 7, pp. 858–866, July 2001.

[41] J. Cong, D. Z. Pan, and P. V. Srinivas, Improved crosstalk modeling for noise constrained interconnection optimization, *Asia and South Pacific Design Automation Conference*, pp. 373–378, 2001.

[42] M. R. Becer et al., Analysis of noise avoidance techniques in DSM interconnects using a complete crosstalk noise model, *IEEE Proceedings of Design, Automation and Test in Europe Conference and Exhibition*, pp. 456–463, 2002.

[43] D. Sylvester and C. Hu, Analytical modeling and characterization of deep-submicrometer interconnects, *Proc. IEEE*, Vol. 89, No. 5, pp. 634–664, May 2001.

[44] L. H. Chen and M. Marek-Sakowska, Closed-form crosstalk noise metrics for physical design applications, *IEEE Proceedings of Design, Automation and Test in Europe Conference and Exhibition*, pp. 812–819, 2002.

[45] A. Odabasioglu, M. Celik, and L. T. Pileggi, PRIMA: passive reduced-order interconnect macromodeling algorithm, *Proceedings of the International Conference on Computer Aided Design*, pp. 58–65, Nov. 1997.

[46] Y. I. Ismail, E. G. Friedman, and J. L. Neves, Equivalent Elmore delay for *RLC* trees, *IEEE Trans. Comput. Aided Des. Integrated Circuits Syst.*, Vol. 19, No. 1, pp. 83–97, Jan. 2000.

[47] A. B. Kahng and S. Muddu, An analytical delay model for *RLC* interconnects, *IEEE Trans. Comput. Aided Des. Integrated Circuits Syst.*, Vol. 16, No. 12, pp. 1507–1514, Dec. 1997.

[48] Y.-C. Lu, M. Celik, T. Young, and L. T. Pileggi, Min/max on-chip inductance models and delay metrics, *Proceedings of the Design Automation Conference*, pp. 341–346, 2001.

[49] X. Huang et al., *RLC* signal integrity analysis of high-speed global interconnect, *Technical Digest, International Electron Devices Meeting*, pp. 731–734, Dec. 2000.

[50] Y. Massoud, S. Majors, T. Bustami, and J. White, Layout techniques for minimizing on-chip interconnect self inductance, *Proceedings of the Design Automation Conference*, pp. 566–571, 1998.

[51] T. Sato et al., Bidirectional closed-form transformation between on-chip coupling noise waveforms and interconnect delay-change curves, *IEEE Trans. Comput. Aided Des. Integrated Circuits Syst.*, Vol. 22, No. 5, pp. 560–572, May 2003.

[52] R. Arunachalam, K. Rajagopal, and L. T. Pileggi, TACO: timing analysis with coupling, *Proceedings of the Design Automation Conference*, pp. 266–269, 2000.

[53] P. Chen, D. A. Kirkpatrick, and K. Keutzer, Switching window computation for static timing analysis in presence of crosstalk noise, *Proceedings of the International Conference on Computer Aided Design*, pp. 331–337, Nov. 2000.

[54] B. Thudi and D. Blaauw, Non-iterative switching window computation for delay-noise, *Proceedings of the Design Automation Conference*, pp. 390–395, 2003.

[55] Y. Sasaki and G. D. Micheli, Crosstalk delay analysis using relative window method, *IEEE International ASIC/SoC Conference*, pp. 9–13, Sept. 1999.

[56] Y. Sasaki and K. Yano, Multi-aggressor relative window method for timing analysis including crosstalk delay degradation, *Proceedings of Custom Integrated Circuits Design Conference*, pp. 495–498, 2000.

[57] J. D. Ma and L. He, Toward global routing with *RLC* crosstalk constraints, *IEEE/ACM Design Automation Conference*, June 2002, pp. 669–672.

[58] M. R. Becer, D. Blaauw, V. Zolotov, R. Panda, and I. N. Hajj, Analysis of noise avoidance techniques in DSM interconnects using a complete crosstalk noise model, *Proceedings of the Design, Automation and Test in Europe Conference and Exhibition*, pp. 456–463, 2002.

[59] D. Sylvester and K. Keutzer, System-level performance modeling with BACPAC: Berkeley advanced chip performance calculator, *Proc. SLIP*, pp. 109–114, 1999; *http://www.eecs.umich.edu/~dennis/bacpac/.*

[60] I. Dobbelaere, M. Horowitz, and A. El Gamal, Regenerative feedback repeaters for programmable interconnections, *IEEE J. Solid-State Circuits*, Vol. 30, No. 11, pp. 1246–1253, Nov. 1995.

[61] A. Nalamalpu, S. Srinivasan, and W. P. Burleson, Boosters for driving long onchip interconnects: design issues, interconnect synthesis, and comparison with repeaters, *IEEE Trans. Comput. Aided Des. Integrated Circuits Syst.*, Vol. 21, No. 1, pp. 50–62, Jan. 2002.

[62] R. Colwell and R. L. Steck, A 0.6 μm BiCMOS processor with dynamic execution, *IEEE International Solid-State Circuits Conference*, pp. 176–177, 1995.

[63] C. J. Alpert, A. Devgan and S. T. Quay, Buffer insertion for noise and delay optimization, *35th IEEE/ACM Design Automation Conference*, pp. 362–367, 1998.

第 9 章　超低功耗电路设计

9.1　引言

过去的 30 年中，促使器件尺寸持续缩小的技术进步使设计者可以不断采用更快的器件、更高的集成容量和更低的动态功耗。所有这些提升使得这一期间微处理器的性能提升了 5 个数量级。然而，随着未来应用对性能要求的持续增长，小于 90nm 节点的工艺缩小变得更加困难。这种发展趋势主要的障碍之一是由性能驱动的尺寸缩小和集成度的提高使业已过大的芯片功耗问题更加严重。在当前的缩小趋势下，每一代工艺进步带来每个节点电容减少 30%、电气节点集成度增加两倍、芯片尺寸增加 14%、电源电压减小 15%、并且频率提升两倍。其结果是，根据工业界提供的数据，每两年 CPU 动态功耗就增加近 2.7 倍[1]。在学术领域，图 9-1a 所示为 1980 ~ 2000 年期间在 ISSCC 上发表的关于前沿处理器的调查，数据表明，每三年功耗增加 1.4 倍[2]。另外，缩小过程中，随着 V_{th} 的减小，导致每代工艺的过泄漏电流增加了 3 ~ 5 倍[3]。根据 ITRS 数据，随着尺寸缩小，估计的泄漏功耗增加趋势如图 9-1b 所示。

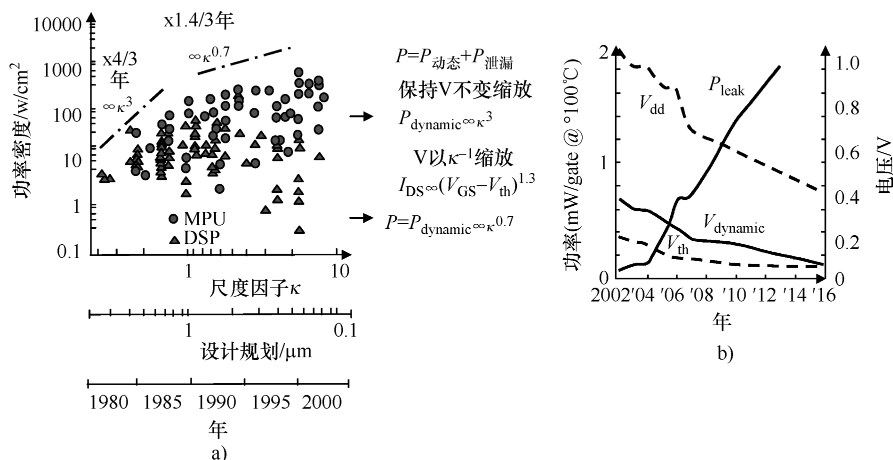


图 9-1　由于器件尺寸缩小引起的动态和泄漏功耗的增加
a）1980—2000 年期间在 ISSCC 上发表的处理器功耗数据　b）ITRS 预测的每器件电压和功耗随着器件尺寸缩小的变化趋势（来源于参考文献［2］）

高功耗将导致台式计算机和服务器的性能与可靠性发生退化。随着微结构的复杂度日益增加，时钟频率和晶片尺寸不断提升，下一代多处理器服务器外壳将可能需要液体冷却或制冷硬件。这种转变可能会使不断上升的性价比上式曲线上呈现一个开始下降的转折点[74]。从性能角度考虑，过多的泄漏功耗减小了如笔记本电脑、手机和 PDA（Personal Digital Assistant，个人数字助理，即掌上电脑）等电池供电设备应用的工作时间。功耗限制已成为 VLSI 发展的一个主要制约因素。因此低功耗设计变成了关键的挑战，特别是对未来低于 90nm 的技术节点，其有效氧化层厚度（EOT）被设定为 1 ~ 1.6nm 范围[35]。采用这种薄 EOT，栅氧隧穿泄漏和栅感应漏泄漏（Gate-induced Drain Leakage，GIDL）变得更加重要并且达到了与亚阈区电流相当的水平。当前，130nm 工艺的待机泄漏一般小于总电流（即待机加动态）的 10%。对 90nm 技术节点，这个比例将增加到 30% 或更大，而到 65nm 时预计会更大。

面对工艺技术进步同时伴随有功率密度持续增大的发展趋势，系统功耗最小化的任务应该包括各个设计级别的优化。从硬件结构、软件操作系统到物理级电路设计，在设计的各个级别中都存在着大量的降低功耗的机会。在采用电池供电的超低功耗应用中，不同设计级别之间针对功耗的各种设计技术的协同是使功耗最小的关键。其他对功耗预算不太严格的计算密集型设计也会从这些技术中部分受益，实现优化的工作效率。因为功耗已成为纳米 CMOS 设计中性能提高的制约因素，采用低功耗技术也将促使很快推出更低功耗的处理器和服务器。

在应用这些功耗减小技术时，从设计过程中的方方面面考虑，可以将这些技术分为两类，设计阶段的技术和运行阶段的技术。在设计阶段完成并确定设计阶段技术的优化，优化全局功耗，运行阶段的技术针对不同阶段的工作负荷，对设计采用不同的实时控制。泄漏抑制技术主要适用于运行阶段，因为它们仅在系统空闲时存在。9.2 节和 9.3 节将概括介绍在系统设计的各个级别中目前运用的设计阶段和运行阶段的功耗控制方法，重点在于电路级的逻辑和存储器设计技术。9.4 节介绍低功耗设计的技术革新。适用于 90nm 以后技术的未来超低功耗设计技术展望将在 9.5 节讨论。

9.2 设计阶段的低功耗技术

9.2.1 系统级和结构级设计阶段的低功耗技术

在系统级，功耗减小技术的目标是使不必要的动作最小化。系统分割技术利用局部性原理把系统或算法分割为局部的簇，这样会使局部总线更短，并且使高容性全局总线的活动降低。在芯片装配时，需要进行优化以限制较宽总线的长

度，只考虑采用更窄的长总线。必须调整布局计划采用信号位数较低的总线来减小宽总线的长度。其他技术包括事件驱动的设计方法、最小化的数据传输、考虑功耗的介质访问协议和网络布线等[4,5]。

在结构级，采用并行硬件实现的设计可以减小电源电压和时钟频率而不会影响系统流量。优化的分层存储器系统可以减小存储器读取次数并应用缓存方案从存储器中存取数据的局部化受益。对于考虑功耗的编译器在编码大小和速度之间进行优化折衷可以减小功耗。功耗经济的 I/O 互连设计减小了总线开关电容并采用数据编码使总线状态切换最小[4]。

9.2.2　电路级设计阶段的低功耗技术

在电路级，有许多技术可以被用来建立功耗优化的电路。

1. 在设计阶段使用层叠效应

如果将两个关态晶体管层叠，由于栅—源、体偏置和漏源电压同时得到减小，使亚阈泄漏电流比单个关态晶体管的小得多。这种层叠效应已经被广泛地应用到各种泄漏减小技术中。这些方法中，大部分使用复杂的低泄漏输入向量、栅调整[45]和插入串联晶体管[44]将待机电路转换为层叠结构等方法，实现运行阶段的待机控制。详细内容将在 9.3.2 节介绍。在设计阶段，基于强制层叠技术[27]，将一个非层叠器件变成两个器件的层叠而不影响输入负载（见图 9-2）。采用这种方法，以一定的延迟增加为代价使层叠驱使逻辑门的泄漏减小了 9 倍，其效果与双 V_{th} 技术相似，但不会像采用第二种 V_{th} 那样增加工艺复杂性。可以在非关键路径应用层叠驱使技术，能够同时减小待机泄漏和有源状态下的泄漏，而会对采用正常门设计的关键路径的速度产生影响。同样的工作结果表明，这种用于减小泄漏的层叠技术将会随着工艺技术的进步和器件尺寸的减小而得到改善，使得在未来的技术发展中采用层叠效应的泄漏控制技术更加有效[27]

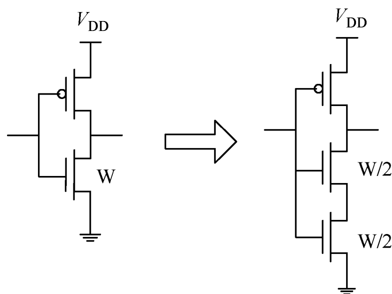


图 9-2　采用层叠技术的双层叠 NMOS
（来源于参考文献［27］）

2. 输入序列重排

输入序列重排是可以优化电路延迟和容性功耗的门级技术。合适的输入序列重排使内部节点的开关活动最小并减小了有源功耗。人们基于电路结构、内部节点电容、（逻辑 1）信号概率和翻转概率的解析模型，提出了不同的算法，实现优化的输入序列[47-49]，总结了输入序列重排的一般规则。这些规则中最得到普遍认同的一点是将转换（开关）率最高的信号放置到最靠近输出端点的地

方[48,50]，这样可以使电源轨的连接活动最小，同时导致性能的优化。这一领域的另一项研究进一步考虑了扇入、扇出门和序列重排门总功耗的优化[51]。所有这些输入序列重排算法可以使平均功耗减小 3.6% ~12%。与其他低功耗技术相比，输入序列重排技术节省的功耗有限；然而，它不要求额外的器件和结构调整，因此可以很容易与其他低功耗技术一起使用。这一特点加速了它的应用。

3. 优化确定晶体管尺寸

在设计时为了得到期望的功耗、时延和面积之间的折衷，晶体管尺寸的确定是一个重点。使用 TILOS[38] 和 EinsTuner[39] 等优化工具，可以广泛地实现尺寸优化。对于给定的时延约束，采用这些工具可以得到使电路总功耗最小化的近似解。作为尺寸优化的第一个综合器，TILOS 在多项式设计优化中[38]采用的是一个简单的 *RC* 时延模型，可以处理多达 25 万个晶体管的电路。在各种高性能芯片设计中使用 TILOS 可以减小 40% ~50% 的功耗[7]。十几年后，IBM 研究开发了 EinsTuner 工具，对于与沟道相连的元器件进行精确的仿真，改善了 TILOS 中的时延模型，并且实现了基于梯度的非线性优化。其结果是 EinsTuner 以更高的精度实现了更好的解决，但是其付出的代价是降低了处理大规模集成电路的能力(对于由 2796 个晶体管构成的加法器电路，包含了超过 5600 个变量和超过 5600 个约束条件，计算过程则花费了三天时间)[39]。另外，通过考虑短路功耗[40]和上升/下降时间延迟单元[41]等，许多其他工作也使尺寸优化得到进一步改善。

4. 应用多电源和多阈值电压

电源 V_{DD}和阈值电压 V_{th}是使有源功耗、泄漏功耗和电路性能之间平衡得到优化的关键因素。在运行阶段，可以动态改变 V_{DD}和 V_{th}，以提高不同负载点的系统功率效率，详细内容在 9.3.1 节介绍。在设计阶段则需要做大量的工作以解决高速低功耗设计中的 V_{DD}和 V_{th}优化问题。目前已推导得到考虑短沟道效应和 V_{th}变化的解析形式方程[58]。以 V_{DD}、V_{th}和尺寸作为变量进行的灵敏度平衡分析结果表明，在延迟增大 20% 的代价下，可获得 40% ~70% 的功耗节省[9]。另一方面，由于一个电路的全局性能一般只受到几条关键路径的制约，而整个电路的路径时延分布范围很广，这也促使了多 V_{DD}和多 V_{th}的使用[54]。如图 9-3a 所示，双 V_{th}技术可以用低 V_{th}器件来对关键路径加速而在非关键路径采用高 V_{th}来抑制泄漏。图 9-3b 显示了双 V_{th}优化对路径延迟分布的影响，其目的是实现路径延迟之间的平衡并加速关键路径。该技术已在电路中得到广泛应用[55,56]，通过在路径平衡中结合多 V_{DD}和晶体管尺寸缩放设计技术，可以使优化空间得到进一步扩展。通过这种优化技术的研究，目前已提出了几种算法，可以在非关键路径上选择能够采用高 V_{th}值而不影响系统整体性能的晶体管（可以通过将一条非关键路径转换为关键路径保证系统整体性能)[57]。另一项工作结果表明，对于有效的设计，V_{DD}、

V_{th}和尺寸的确定无需超过三个离散值[8]。

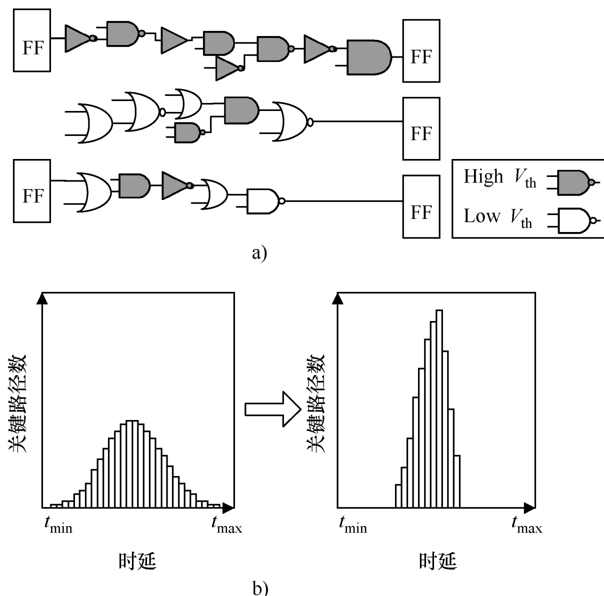


图 9-3　采用双 V_{th}技术的高速低泄漏设计

a）关键路径上的晶体管应用低 V_{th}值　b）双 V_{th}优化前后的路径延迟分布

5. 器件沟道长度采用非最小值

设计中通常都采用工艺允许的最小沟道长度。到 130nm 技术节点，这一直是传统的设计惯例，但是现在由于泄漏电流的显著增加，设计者不得不将器件分成最小的和非最小的沟道长度以减小泄漏电流。沟道长度的增加对设计电路的总待机电流有很大的影响，特别是如果它被广泛地应用到大量的器件中时。图 9-4 所示为 100nm 工艺中源—漏泄漏电流与沟道长度的关系。所有的泄漏电流值均参照沟道长度为 15μm 的器件进行了归一化处理。将沟道长度从 100nm 增大到 150nm，可以实现约 60% 的泄漏减小。与双 V_{th}情况相似，可以在非关键速度路径上使用非最小化沟道长度以平衡系统延迟分布。采用非最小化沟道长度的另一个优点是可以减小工艺波动对这些路径的影响，因为在更长沟道器件中，沟道长度波动的相对值减小了。另外，使用多阈值器件方法将因为要包括附加的掩膜和更多的工艺步骤而增大工艺费用，而器件沟道长度采用非最小

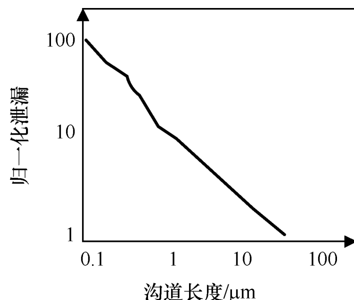


图 9-4　归一化 I_{ds}泄漏与沟道长度的关系

值的设计方法可以提供成本更低的方案。

6. 低功耗标准单元库和定制库产生

采用由功耗最小的元件形成的标准单元库极大地方便了低功耗系统的实现。低功耗标准单元库中的单元采用能量高效的逻辑类型和定制尺寸形成，并针对不同的规范要求，提供有适用于不同阈值电压的版本。结果表明，与采用通用标准单元库得到的设计相比，使用适当策略，由逻辑综合器基于低功耗库产生的设计具有更好的性能，并使功耗得到改善[10]。采用定制库的方式[52]则可以进一步克服定制低功耗标准单元库尺寸固定的问题，给设计提供了更大的灵活性。基于按需求产生的 ASIC 设计方法如图 9-5 所示。其中根据性能评估的结果产生一个经修正的单元库，并将其用到基于单元的设计工具中。该设计流程的特点是在版图设计后再进一步确定晶体管尺寸，即基于从初始版图中提取的信息来缩小单元尺寸，从而优化单元库。采用这种方法，可以消除采用通常固定库设计中存在的面积和功耗冗余问题，形成一个完全优化的物理设计。据报道，采用定制库的方法，可以使电路功耗的减小最大可达 77%，平均为 65%，并且不会增加时延[53]。

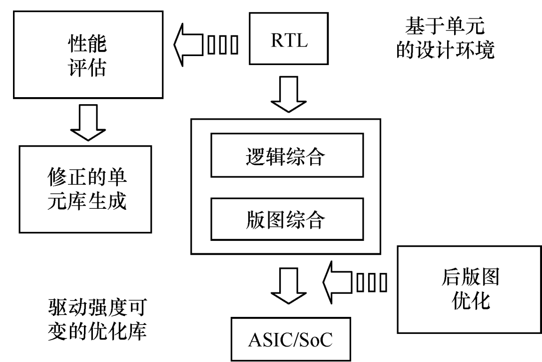


图 9-5　使用定制库的 ASIC 设计方法（来源于参考文献［52］）

7. 减小互连功耗

互连，包括片上互连和封装引线，已成为功耗的主要来源。随着金属互连层数的增加互连尺寸的缩小，目前设计中金属互连电容已占据芯片总电容的 70%[2]。另外，由于动态功耗为（$C \cdot V_{dd}^2 \times$频率），因此芯片工作频率的快速增加进一步恶化了互连系统中的动态功耗。注意，互连感抗在电压翻转时并没有直接消耗功率。基于互连功能划分的下述三种互连是决定功耗的主要因素：片上信号线、I/O 系统的互连和时钟分布网络。为了减小他们的功耗，业已从技术和设计两方面提出了许多创新技术。减小互连功耗的一种普遍方法是采用低电压摆幅。这个技术被广泛应用到 I/O 系统中，如低压差分信号（Low-Voltage Differential

Signaling，LVDS）。LVDS 不仅节省功耗而且提高了 I/O 信号传输速度。然而，随着纳米时代信号耦合噪声的急剧增加，信号完整性考虑和设计开销制约了这一方法在片上信号发送和时钟网络中的应用。业已表明，对全局信号互连，通过改变总线位置或者对翻转方式进行编码，将使最坏情况下的耦合电容被最小化，因此采用总线改变和编码技术可以减小耦合电容上的功耗。另一种考虑功耗影响的互连设计方法是引入非正交的全局层，减小总的信号互连长度。例如，在 X 结构中[97]，可以采用 45°的版图。这种情况下，可以节省约 20% 的总互连长度，进而可以减小相同比例的互连功耗[97]。

9.2.3　设计阶段的存储器技术

单片随机存取存储器（RAM）和嵌入式 RAM 具有高密度、低功耗和低成本的特点，为不同电子系统的改善起到重要作用。现今，微处理器设计中融合了大量的存储器单元，而存储器将消耗系统功耗预算的很大部分。例如，Alpha 21246 功耗预算的 30% 和 StrongARM 功耗预算的 60% 都用于缓存和存储器结构[60]。对采用电池供电并且具有低占空比的设备，存储器漏电流功耗甚至可以占据系统总功耗的主要部分并且决定电池的寿命。各种应用中对系统功耗优化的需求，使得低功耗存储器的设计成为一个发展迅速、硕果显著的领域。9.2.3 和 9.3.3 节将分别介绍低功耗 SRAM 和 DRAM 的设计技术和运行技术[6,31]。

1. 低功耗 SRAM 设计

（1）按块划分的字线和位线的局部激活。局部激活是减小 SRAM 中大负载的字线和位线充电电容的极为有效的方式。只要将存储器阵列简单地分割为子模块[14]，就可以极大地减小字线和位线的负载。然而，由于需要附加的控制逻辑和布线，使得这个技术有极大的局限性。其他技术则保持了存储器阵列的完整性，而重点在于译码逻辑的重构。如图 9-6a 所示，分割字线（Divided-Word-Line，DWL）方式[15]采用两级分层行解码器结构。在每次访问存储器期间，只有一条子字线被激活，这样，相比未分割的主字线模式，一般只有 10% ~25% 的电容引入电路，这就就相应地减小了功耗和字线时延。DWL 方式在 1MB 以及更高密度的 SRAM 中已被广泛地使用[14]。为了进一步减小电容，还可以采用将 DWL 和多行解码器及三级分层行解码方案结合起来的方法。还有一种单个位线交叉点单元激活（Single-bit-line Cross-point cell Activation，SCPA）结构[59]的方式，即通过单个单元激活来实现位线电流最小化。如图 9-6b 所示，存储器存取仅激活 X 和 Y 地址控制的交叉点上的那个 SRAM 单元。据报道，与传统的 DWL 结构相比，采用 SCPA 实现的 16MB SRAM 可以减小 36% 的激活电流和 10% 的面积。

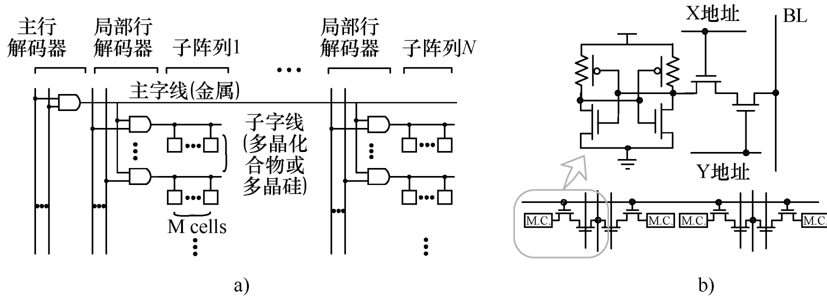


图 9-6　按块分割字线的局部激活方案

a）DWL 结构　b）SCPA 结构使用的 SRAM 单元

（a 来源于参考文献［15］；b 来源于参考文献［59］）

（2）脉冲操作。脉冲字线（Pulsed Word Line，PWL）操作可以将工作时的占空比降低到读和写操作所需的最小时间[61]，因此就减小了存储器存取期间的动态功耗。图 9-7a 显示的是 PWL 的部分电路图和时序图。该电路中的地址转换检测（Address Transition Detection，ATD）[6] 单元用来检测地址和检测信号转换，并产生脉冲信号。电路图中所示的 XD 脉冲在 ATD 上升时形成，并通过 X 解码器和读出放大器控制字线操作。ATD 脉冲产生电路如图 9-7b 所示。这个脉冲电路也可以用于高容性负载预解码线、写总线、位线和读出电路[16-20]。

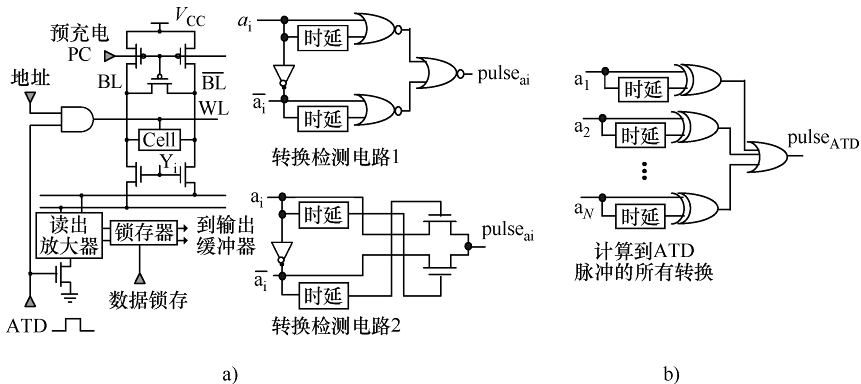


图 9-7　PWL 操作

a）部分电路和时序图　b）ATD 脉冲产生电路

（a 来源于参考文献［61］；b 来源于参考文献［6］）

（3）降低 V_{DD}时的单元驱动方案。减小电源电压能够有效地抑制 SRAM 的动态和静态功耗，因此，在过去的十年间，出现了许多工作电压低于 1V 的低功耗 SRAM。为了在电源电压为0.5～0.8V 时实现 100MHz 操作，设计中采用了各种

各样的单元驱动形式，如驱动源极互连（Driving Source Line，DSL）[11]，负字线驱动（Negative Word-line Driving，NWD）[12]以及升压接地补偿数据存储（Boosted Offset-Grounded data Storage，BOGS）[13]。图 9-8 显示的是 DSL 和 NWD 的单元电路图以及工作波形。DSL 方式是在读取操作期间，将交叉耦合反相器的源极互连连接到负电压 V_{BB}上，而在写操作期间源极互连悬空。这样，单元的读取时间由于栅源电压增加和晶体管源——衬底结的正向偏置的提升而得到了改善。由于交叉耦合反相器对中的 NMOS 晶体管处于未激活态，写周期也同样得到了改善。NWD 方式使用了截止栅压为负的低 V_{th}导通管（Q_1和 Q_2），和提升了栅压（$V_{CH} > V_{CC}$）的高 V_{th}交叉耦合反相器对来实现存取时间的改善和待机泄漏电流的减小。通过对栅源偏置和 V_{th}的控制，与常规的单元相比，DSL 和 NWD 方式在低于 1V 的电源电压下提高了存储器的工作速度，并且抑制了静态待机泄漏电流。然而，这些方案的应用中包括了几种系统开销，如 DSL 方式下产生负电源电压的电荷泵运行效率低和 NWD 方式下从提升电压增加的存储节点到位线的高

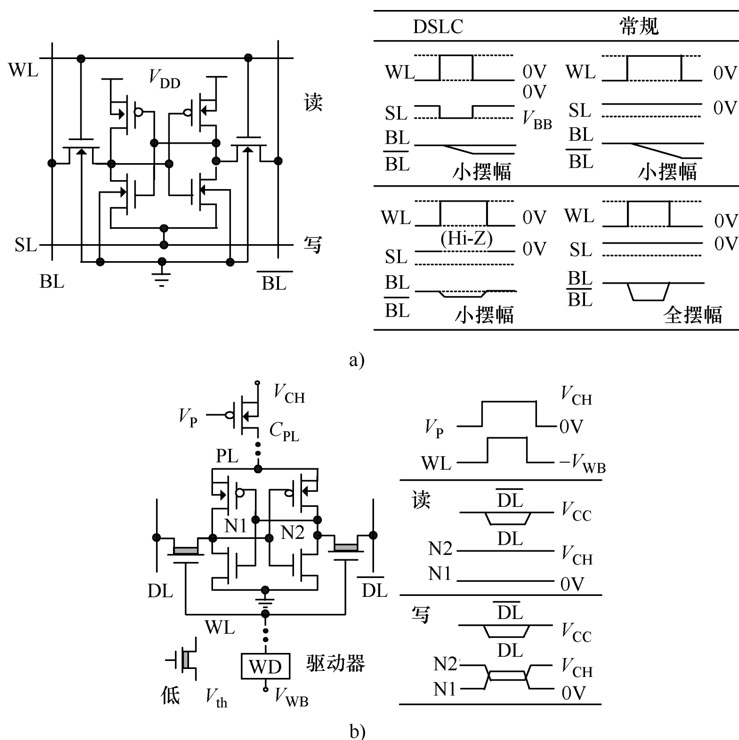


图 9-8　DSL 和 NWD 方案

a）DSL 单元和读/写周期时序图　b）NWD 单元和读/写周期时序图

（a 来源于参考文献［11］；b 来源于参考文献［12］）

泄漏流。BOGS 就是为解决这些问题而提出的另外一种单元驱动电路。其中将数据存储节点对的电势从 0.5V/0V 变到 1.3V/0.65V，这样就消除了产生负的源极互连电压的需求。调整位线预充电和字线驱动之间提升的电势，可以避免从电压增加的存储节点到位线的泄漏。该方法同样也使用了循环充电的方式以节省源线电压控制电路的功耗。

（4）低功耗读出放大器设计。I/O 线上的读出放大器一般消耗 1～5mA 的直流电流[6]。为了获得更高的数据流量，高速处理器会增加其 I/O 线的数目，这样读出放大器上的功耗将成为芯片总功耗中一个很大的组成部分。如图 9-7 所示，通过仅在脉冲高电平期间使电路工作，这种脉冲操作方案可以有效地减小读出放大器的功耗。图 9-9a 所示为一种 1989 年提出的锁存型 PMOS 交叉耦合读出放大器设计[62]。与通常的电流镜运放对相比，这种设计使读出时延减小了 50%，并且使全输出摆幅时的电流减小了 80%。为了稳态工作，需要精确的时序来控制均衡器，从而使该放大器输出达到平衡。图 9-9b 所示为另一种高速读出放大器的设计[63]。该放大器读出的是位线电流差异而不是电压差。采用这种设计，可检测的数据线电压摆幅降低到小于 30mV。与常规的要求电压摆幅为 100～300mV 的电压型读出放大器相比，在延迟时间固定为 1.2ns 的情况下，这种电流型读出放大器设计节省了 60% 的功耗[64]。因为这种设计中的位线电压保

图 9-9 高速低功耗读出放大器设计

a）PMOS 交叉耦合放大器设计 b）电流读出放大器设计

（a 来源于参考文献［62］；b 来源于参考文献［63］）

持相等，读出放大器本身拥有均衡的功能，因此简化了工作的时序控制。

2. 低功耗 DRAM 设计

在过去的 10 年中，电路技术的持续发展，使 DRAM 在容量不变的前提下，功耗减小了两到三个数量级。与 SRAM 的情况相似，DRAM 中减小的动态电流有助于实现低功耗、低结温和低成本封装。充电电容和工作电压的减小已成为控制 DRAM 动态功耗的主要技术。同时，亚阈区电流抑制措施，如待机负栅源偏置等，将是未来电池供电的 DRAM 系统中必不可少的[6]方法。

作为 VLSI 存储器系统中的关键部件，DRAM 和 SRAM 在工作、结构和功耗产生原因等方面都是相似的。因此，它们可以共享许多相似的降低功耗技术。因为本章节已对低功耗 SRAM 设计进行了讨论，因此，下面只对 DRAM 设计阶段和运行阶段的功耗控制进行简要叙述，而着重于专门针对 DRAM 结构的设计技术。

（1）充电电容的减小和刷新时间的延长。与 SRAM 中的 DWL[15] 和 SCPA[59] 相似，分块的数据和字线的局部激活技术可以用来减小充电电容。这就降低了动态功耗，并提高了存储器访问时的信噪比。图 9-10a 和图9-10b 分别所示为采用 DRAM 数据线和字线局部激活的电路示意图[87,88]。在这些方法中，数据和字线被分割成多个部分。这些子块由附加的控制逻辑激活，例如部分数据线激活中的 Y 解码器和部分字线激活中的行选择线 RX。共享的 I/O、读出放大器和解码逻辑有助于减小控制电路的系统开销[87]。配合部分激活方案一起使用的另一种静态电流减小技术是刷新时间的延长[87]。通过灵活地控制数据线的子片段，可以延长存储器刷新时间而不会影响正常的操作。这是通过在刷新周期控制数倍于普通周期中同时激活的阵列数实现的。自刷新时间的延长使得刷新电流和 DRAM 静态功耗均得到减小。

（2）工作电压的减小。受面积尺寸和低功耗需求的驱使，DRAM 电源电压已经从 20 年前的 12V 减小到现在接近 1V 的水平。然而，进一步缩小到 sub-1V 范围则面临着巨大的挑战，因为此时将会产生工作速度的衰退和 V_{th} 的减小引起泄漏功耗的恶化。克服这些困难的关键在于高速读出放大器与存储器操作的设计以及有效的亚阈区泄漏抑制策略，这些内容将在后面的运行阶段的存储器低功耗技术部分进行介绍。另外，对数据线采用 VDD/2 预充电的方式[89]，在减小电压摆幅的情况下可将数据线功耗减小一半。同时，存储和预充电时引起的较大尖刺电流也被减半，从而可以实现低噪声操作。最后，在各种各样的存储器功耗控制技术中必不可少的一项技术就是片上降压转换器，可以产生所需的不同电压，如（VDD/2）数据线操作方案中的预充电电压。这些转换器在负载电流快速变化的情况下提供稳定且精确的输出电压[6]。

a)

b)

图 9-10　DRAM 功耗减小的部分激活方案

a）多次分割数据线的部分激活　b）多次分割字线的部分激活

（a 来源于参考文献［87］；b 来源于参考文献［88］）

9.3　运行阶段的低功耗技术

9.3.1　运行阶段的系统级和结构级低功耗技术

基于如工作负载这样的实时操作信息，运行阶段的系统级低功耗技术主要用

来优化系统管理策略。这些技术包括与功耗相关的不同动态安排方案（根据估计的执行时间来安排任务[21]）、动态功耗管理（Dynamic Power Management，DPM）（动态地重新配置一个电子系统，以最小数量的有源元件提供要求的服务[22]）、以及通信网络应用中与功耗相关的布线[23]。

在结构级中，如果是通过反馈环路控制 V_{DD} 和系统时钟频率来执行一个特定的任务，采用动态电压和频率缩小技术（Dynamic Voltage and Frequency Scaling，DVS，也称为 DFS 和 DVFS）[24]是广为人知的功耗减小方法。为了进一步减小空闲期间的泄漏，可以采用动态 V_{th} 缩放（Dynamic V_{th} Scaling，DVTS）[25]方法，通过体偏置控制的方式实现阈值电压的自适应调整。衬底正偏（Forward Body Bias，FBB）和定向自适应衬底偏置（Adaptive Body Bias，ABB）已用作为通常衬底反偏（Reverse Body Bias，RBB）控制的增强技术。FBB 具有我们期望的改善晶体管短沟道效应的效果，这样就减小了对特征尺寸变化的敏感度[85]。为了补偿芯片范围内的参数波动效应，提出了芯片内 ABB（WID-ABB）技术。该技术集成有相位检测器并且分别给每个电路模块提供合适的衬底偏置。ABB 和 WID-ABB 的作用效果如图 9-11 所示。只采用 ABB 技术就可以使芯片范围频率分布的标准偏差 σ 减小 7 倍，结合采用 WID-ABB 技术，可以使 σ 再减小三倍，从而在最高的频率窗口下几乎 100% 的芯片都满足要求[86]。V_{DD} 和 V_{th} 跳跃技术与 DVS 技术类似。该技术将 V_{DD} 和 V_{th} 调整为由软件反馈环控制的几种离散值[26]。

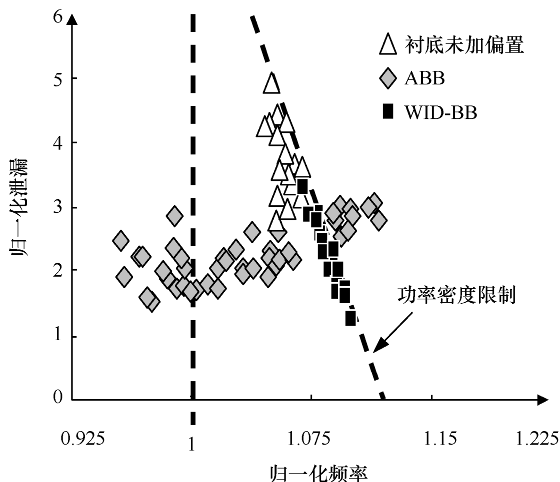


图 9-11　ABB 和 WID-ABB 控制对泄漏与频率分布的影响（来源于参考文献［86］）

FBB 和 ABB 技术不是没有问题。衬底噪声将对电路性能产生影响，除非对衬底偏置电源进行很好的去耦处理，使其像一个独立的电源。但是去耦需要占用布线资源，同时去耦电容还会增大芯片面积。如果 FBB 将体电压抬升得太高，

使晶体管的结正向偏置，还会有发生闩锁的危险。FBB 也会增加结电容并增加芯片的动态功耗。自适应衬底负偏置是低功耗设计愿意采用的方式。衬底负偏置减小了结电容和亚阈区泄漏，因此能改善动态功耗和待机功耗。

为了使衬底偏置能起到作用，必须使晶体管在较高的体效应下才会开启，这样负的衬底偏置将会提升 V_{th}，从而减小泄漏电流。栅感应漏泄漏（GIDL）在 90nm 节点和尺寸更小的工艺中影响将会明显。GIDL 改变了器件的亚阈区曲线，使得泄漏电流随着栅驱动的减小而增加。衬底偏置加剧了这个效应，使衬底偏置起了负面作用（见图 9-12）。

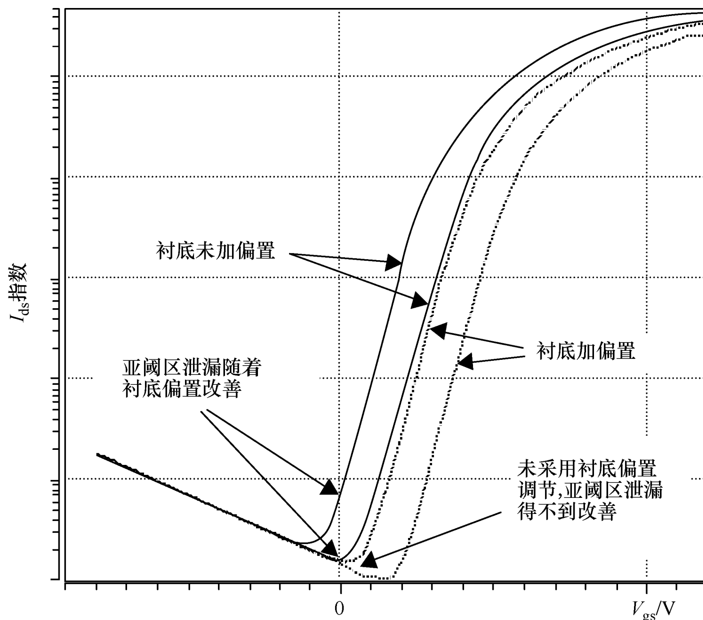


图 9-12　为了使衬底偏置效应发挥作用，晶体管必须进行调整

对于可靠性要求较高的服务器类高性能微处理器，需要进行可靠性老炼。许多情况下，由于电源功率有限，为了防止老炼过程中只有一部分得到老炼，就要限制同时老炼的器件数。这严重限制着老炼的产量，迫使设计者通过牺牲性能的途径来减小功耗，使得同一个烘箱中，器件中更多的部分可以同时得到老炼。如果采用衬底负偏置可以用来减小老炼中的亚阈功耗，这样设计者就无需牺牲性能来方便微处理器的老炼[104]。

时钟分布网络是另一个主要的功耗源，特别是在高性能微处理器设计中。在 72W 600MHz 的 Alpha 处理器中，一半功率消耗在时钟网络上[93]。在解决这个问题的各种方法中，时钟门控[75]是一项重要的针对运行阶段的结构级低功耗技术。该技术通过阻止逻辑模块中的不必要活动以及消除时钟网络中不必要的功耗，使

得与时钟有关的有源功耗得到有效的最小化。虽然只要在电路中插入开关就可以很方便地实现时钟门控，但是给结构级优化带来了许多问题，包括采用最少总互连长度的时钟树构建、时钟的时序约束管理、门控时钟网络不对称的最小化等。除了时钟门控，还提出了新的共振时钟结构，使用耦合行波和驻波振荡器或者基于螺旋电感的共振网格[94-96]。在这些方法中，电磁能量在 *LC* 系统中振荡而不是像 *RC* 那样转化为热，因此，减少了功率损耗。在谐振频率为 1.1GHz 时，可以获得高达 80% 的时钟功率节省[96]。

9.3.2　针对运行阶段的电路级低功耗技术

1. 在运行阶段使用层叠效应

与设计时将单个器件变为层叠器件的强制层叠方法相似[27]，针对运行阶段的各种其他低功耗技术通过将待机模式下的电路转换为层叠结构也能从层叠效应得到好处。考虑到在不同的输入向量作用下电路的泄漏功耗在较大范围变化较宽广[42]，这些技术的目标是通过使待机状态的器件采用低泄漏输入向量的方法来减小门泄漏。为了获得与最小泄漏功耗相对应的输入向量，业已提出了大量的算法，包括置信水平给定时的随机采样[42]、原始估计[43]、以及基于可观察性泄漏测量[44]和布尔网络建模[45]的启发式搜索。对 32 位静态 CMOS Kogg-Stone 加法器，如果在输入端施加一个特定的输入向量，可以实现高达 20% 的泄漏减小[46]。对于具有较大逻辑深度的电路，已使用基于多路复用器插入和栅调整的方法来对内部节点实施控制[45]。如图 9-13a 所示，多路复用器的插入可以实现对内部节点 X 的存取。图中采用一个与门作为多路复用器，因为多路复用器的一个输入电平是固定值。图 9-13b 介绍了两种调整全互补 CMOS 门的方法，使得

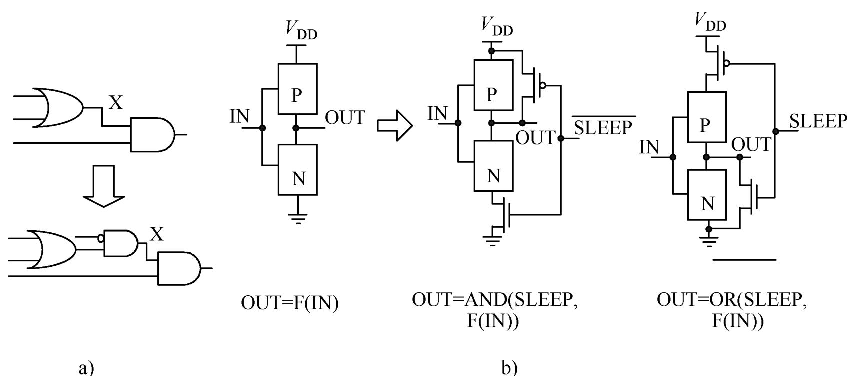


图 9-13　应用输入向量控制电路内部节点的方法

a）插入多路复用器（简化为与门）　b）门调整使输出控制成为可能。（来源于参考文献［45］）

在待机状态下，器件输出端连接到 1 或 0。采用这种方案，层叠效应使调整后的门和它的扇出门的泄漏得到减小。采用这两种方案的应用输入向量控制的基准电路以降低速度和增大面积为代价，节省了 10% ~70% 的泄漏功耗[45]。除了这些针对内部节点的方法，另一种方式是在高泄漏状态下给这些内部门插入一个串联的低 V_{th}开关[44]。在待机期间，这个串联泄漏控制开关切断了那些输入向量控制不能控制的泄漏门。

2. 多阈值 CMOS（Multi-Threshold CMOS，MTCMOS）

在电路级，有代表性的针对运行阶段低功耗技术是多阈值 CMOS（MTCMOS）[64]、可变阈值 CMOS（Variable-Threshold CMOS，VTCMOS）[65]、动态阈值 CMOS（Dynamic-Threshold CMOS，DTCMOS）[66]和其他的衍生技术。这些技术通过在待机模式下插入串联电阻或者增加器件阈值电压 V_{th}的方法来减小待机泄漏电流。如图 9-14a 所示，MTCMOS 通过串联高 V_{th}功率开关的方式关断了低 V_{th}逻辑模块。为了平衡工作时延和面积开销，需要一个尺寸合适的高 V_{th}开关。为此，人们开发出了一种层次化的尺寸确定算法，针对给定的时延约束，使总的硅面积最小[67]。MTCMOS 技术已成为众多低功耗设计中的一项非常有效的技术。然而，当 V_{DD}减小到 1V 以下时，由于高 V_{th}器件的开启电压要求较高的电压，将导致 MTCMOS 的效率不断降低，并且最终会失去作用。对未来的低压工作条件，先后提出了包括超级截止 CMOS（Super Cutoff CMOS，SCCMOS）[68]和栅压升高 MOS（Boosted-Gate MOS，BGMOS）[69]在内的改进结构，使基于功率开关的泄漏抑制方案可以继续发挥效用。SCCMOS 方法是给待机模式下的低 V_{th}开关施加负的栅源偏置电压，而 BGMOS 方法则使用升高的栅源过驱动电压以加速高 V_{th}开关的动作。这两个方法都有效抑制了低 V_{DD}时的泄漏电流，但是以额外的电平设计开销为代价。另外，曲折型超截止 CMOS（Zigzag Super Cutoff CMOS，ZSCMOS）和曲折型栅压升高 CMOS（Zigzag Boosted Gate CMOS，ZBGCMOS）方法作为 SCCMOS和 BGCMOS 的衍生技术也被提出来以改善上电时间[76]。

3. VTCMOS

VTCMOS 技术如图 9-14b 所示，针对不同的工作模式，通过调整工作电路的衬底偏置来得到期望的阈值电压。与 MTCMOS 相比，VTCMOS 要求的面积开销更小，并且衬底中的瞬态电流远小于从电源中抽取的动态电流。因为 VTCMOS 的动态工作不受泄漏控制的影响，因此 VTCMOS 的应用不会受到电源电压缩小的限制。然而，随着技术朝更小的沟道长度发展，衬底偏置控制对 V_{th}的影响变得更弱[70]。衬底反偏置对短沟器件的影响，增加了芯片内 V_{th}的分散性，从而减小了这一技术的有效性[71]。其结果是，使用正向偏置就成为未来实现 VTCMOS 的首选设计选择[6]。

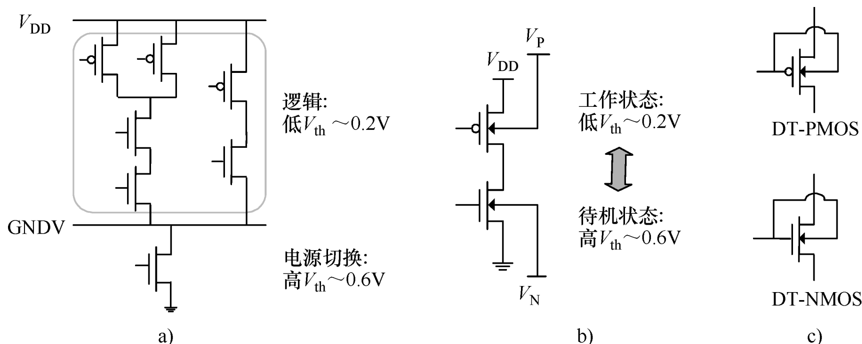


图 9-14　针对运行阶段的电路级低功耗工作方法

a）MTCMOS　b）VTCMOS　c）DTMOS

4. DTCMOS

虽然像 MTCMOS 和 VTCMOS 这类常规低功耗电路技术一直在不断发展以满足未来设计的要求，但 DTCMOS 技术则是在搁置了十年后才获得了电压低于 1V 电路设计的青睐。如图 9-14c 所示，1994 年提出的 DTCMOS 技术采用的是将栅和衬底相连这样一种新颖的 MOSFET 工作方式[72]。这种连接方式使器件的 V_{th} 与栅压密切相关。在动态工作期间，V_{gs} 增加时，V_{th} 会降低，从而能够提供比采用低 V_{DD} 的标准 MOSFET 更大的电流驱动能力。另一方面，空闲时的零 V_{gs} 会使 V_{th} 升高，可以有效地抑制泄漏电流。但是，为了防止过大的衬底电容和衬底电流，DTMOS 的栅压必须小于约一个二极管正向电压（在室温下为 0.7 V），这就限制了这一技术在采用较高 V_{DD} 的设计中的应用。采用 SOI 有助于减小结横截面积，缓解正向偏置的危害。已经提出了几种方法来消除低电压工作的限制，包括使用辅助的 MOSFET 或二极管将衬底-源钳位为正向偏置、或限制它的瞬态效应[73]。当未来的低功耗设计要求低电压工作时，DTCMOS 将成为一项有竞争力的技术。

9.3.3　针对运行阶段的存储器低功耗技术

1. 运行阶段的低功耗 SRAM

（1）通过 SSI 抑制外围电路泄漏。存储器外围电路是由多个相同的电路模块组成的，在待机期间，由于总沟道宽度很大，这些模块将成为多条很严重的泄漏路径。即使在动态工作模式下，除了一小部分被选的模块，这些电路中的其他大部分都处于非激活态。这些特点使采用简单而有效的亚阈区电流控制成为可能。许多降低逻辑电路泄漏电流的技术，如栅——源背偏、衬底——源背偏、多 V_{th} 和功率开关技术都已在存储器外围电路泄漏抑制中得到了广泛使用[91]。作为这些技术应用的一个实例，图 9-15 所示为源极开关阻抗（Switched-Source Impedance，SSI）方案[77]，它关断了空闲期间的控制电路泄漏路径。图 9-13b 中采用

电平固定的输入缓冲器将内部节点驱动到预定的电平。图中的 M_C、M_S和标为阴影的反相器都是高 V_{th}开关。

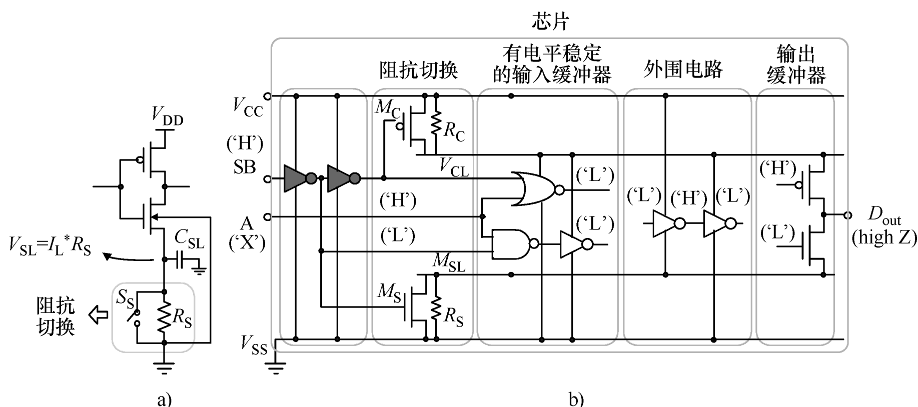


图 9-15 SSI 方案和它在存储器中的应用

a）SSI 电路结构 b）应用到存储器外围电路泄漏抑制中的 SSI（来源于参考文献［77］）

（2）可变阈值的泄漏抑制方案。在电路级泄漏抑制措施中，可调节的体偏置控制具有保持存储在锁存电路中数据的特性，因此被应用在存储器阵列中以减少亚阈区泄漏电流。与 VTCMOS 技术相似，未被选中的存储器单元的衬底电压被反向偏置以获得待机时的高 V_{th}。图 9-16a 所示为是动态泄漏关断（Dynamic Leakage Cutoff，DLC）方案的电路图和时序图[78]。DLC 使未被选中的 SRAM 单元的衬底施加偏置，对 V_{NWELL}大约为 $2V_{DD}$，对 V_{PWELL}约为 $-V_{DD}$。图 9-16b 所示为自动背栅控制 MT-CMOS（Auto-Backgate-Controlled MT-CMOS，ABC-MT-CMOS）技术的配置[79]。在 Q1、Q2 和 Q3 开启而 Q4 关断时，动态工作电压为 1V。在待机模式下，Q4 开启而其他管子关断。ABC-MT-CMOS 的 V_{DD}轨和地轨被二极管 D1 和 D2 钳位。在反向偏置电压 $V_{D1}=V_{D2}=1.15V$ 时，泄漏电流被减小到 20pA/cell。

（3）门控电源和超低待机电源电压方案。在结构级，降低 SRAM 运行期间泄漏电流的技术包括门控关断闲置部分存储器的电源电压或者使得使用频率很低的部分进入睡眠待机模式。这些方法使泄漏功耗的降低与 V_{DD}成二次方关系，并且在编译器级的缓存活跃率分析的帮助下实现了优化的功耗——性能折衷。缓存延迟技术在缓存互连门控中应用了适当的时序策略，以适当的性能损失节省 70% 的泄漏[28]功耗。如图 9-17a 所示，为进一步开发利用控制缓存泄漏电流技术，可以采用将未激活的缓存单元分配到低功耗模状态下，使 V_{dd}得到降低，但是仍然可以保持存储器中的数据[29]。图 9-17b 所示的双轨待机方案是为超低功

图 9-16　动态泄漏关断和自动背栅控制 MT-CMOS

a）DLC 方案的电路图和时序图　b）ABC-MT-CMOS 电路的配置

（a 来源于参考文献［78］　b 来源于参考文献［79］）

耗应用提出的，在待机期间，整个 SRAM 模块被推入到待机电源电压为 300mV 的深度睡眠。在这种超低的数据保持电压下可以减小超过 90% 的泄漏功耗[30]，同时也说明了 0. 13μm 工艺中 300mV 电压就可以实现稳定的数据保持，而数据保持电压（Data Retention Voltage，DRV）与阈值电压和沟道长度的工艺波动之间近似呈现出线性增加的关系。

2. 运行阶段的低功耗 DRAM

（1）外围电路功耗的减小。正如前面提到的，逻辑电路泄漏电流降低技术，如栅——源背偏、衬底——源背偏、多 V_{th} 和功率开关方式等，对存储器外围电

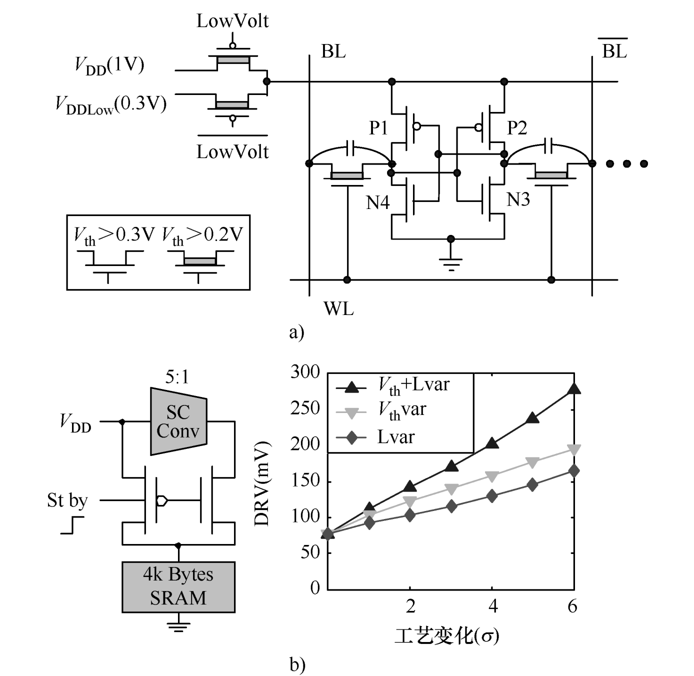


图 9-17　采用超低待机电源电压的 SRAM 泄漏抑制方案

a）处于睡眠的存储器电路　b）双轨 SRAM 待机方案和工艺对 DRV 的影响

（a 来源于参考文献［29］；b 来源于参考文献［30］）

路泄漏抑制来说是很有效的方法。另外，提高片上电压转换器的转换效率并且使它们的待机电流最小对低功耗 DRAM 设计非常重要。这是因为高速低压 DRAM 设计严重依赖于提升各种电流驱动和抑制亚阈区电流的方法。这些方法一般要求多个电平，例如背偏置、基准和预充电电压[6,90,91]。

（2）刷新时间延长和循环充电。如 9.2.3 节提到的，延长刷新时间间隔可减小刷新电流。该方法通过部分数据线激活技术来实现，可为阵列操作提供灵活的控制[87]。循环充电是在刷新期间减小容性数据线功耗的另一种方法。其中，在一个阵列中使用并且在传统情况下每个周期泄放的电荷被转移到另一阵列得到再利用[90]。

（3）DRAM 单元的栅——源补偿驱动方案。提升的读出地（Boosted Sense Ground，BSG）和负字线（Negative Word-Line，NWL）技术作为 DRAM 单元中应用的栅——源补偿驱动方案已广为人知。这两种技术都通过动态的改变 DRAM 单元中晶体管的 V_{th}来得到期望的动态驱动电流和较小的亚阈区电流。如图 9-18 所示，在待机期间，BSG 将未选中单元的数据线电压抬高了 ΔV_{DL}，而 NWL 将栅电压减小了 ΔV_{WL}。在待机状态下增加的 V_{th}可以抑制单元的泄漏电流并使 DRAM

单元中低 V_{th}晶体管的使用成为可能。在上述对比中，常规方案的 V_{th}取为 1V 而 BSG 和 NWL 则为 0.5V。NWL 中 ΔV_{WL}的产生相对容易些，因为与 BSG 中数据线电流沉相比，NWL 中有字线放电电流。但是，这两种方法都增加了未选中单元的栅氧应力[92]。

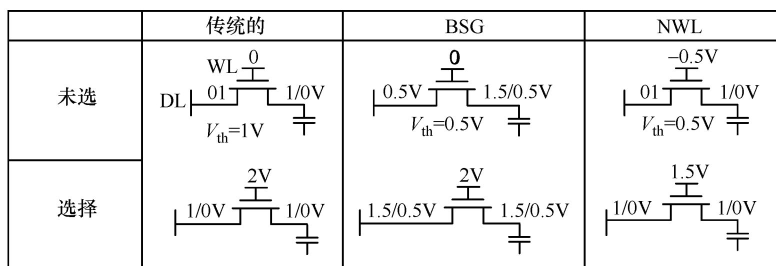


图 9-18　DRAM 单元驱动方案的对比（假定 V_{th0} = 1V，并且存储电压为 1V）（来源于参考文献［92］）

9.4　低功耗设计的技术革新

对未来的设计来说，技术进步导致的尺寸缩小加剧了功率泄漏问题，但是改进的 CMOS 技术也给设计者提供了实现低功耗的工艺途径，包括高 V_{th}的选择、厚栅氧和多阱的选用，这也方便了具有自适应性特点的衬底偏置方法的使用。随着对传统 CMOS 的突破，技术革新带来了许多新颖的器件和制造工艺，包括 SOI、双栅器件和应变 Si。新的组装技术，如系统级封装（System-in-a-Package，SiP），有助于减小封装和板级电容并实现低功耗系统集成。

9.4.1　新颖的器件技术

晶体管尺寸缩小的主要困难在于关态泄漏的控制。为了解决这个问题，人们提出了多种先进的器件结构，例如全耗尽 SOI（例如超薄衬底器件[100]）和双栅结构（例如 FinFET[101]）。其中，FinFET 的工艺流程和版图结构与常规 MOSFET 相似，同时具有优良的可缩放性，因此被认为是 ~10nm 栅长器件技术的最佳选择。FinFET 的结构如图 9-19a 所示。它一般使用 SOI 结构的衬底制造，栅跨在一个鱼鳍状的衬底上，在鳍的两侧形成两个自对准沟道。鳍的顶部一般覆盖着硬掩膜并且不是沟道的部分。鳍的高度（H_{in}）确定了器件的宽度。采用多个鳍的组合就可以实现不同宽度的器件。由于使用双栅

控制很薄的沟道，从而有效地抑制了器件的短沟道效应。FinFET 器件的其他优点包括：由于本征掺杂沟道中的载流子迁移率更高，因此电流更大；另外由于同时消除了耗尽电容和结电容，因此电容更小。这样，由于双栅特性抑制了亚阈区泄漏，FinFET 的待机功耗减小。而且，因为 FinFET 的驱动电流比常规体硅器件更高，就可以减小电源电压，与体硅 CMOS 的同一性能匹配，这样也可以极大地减小了电路的有源功耗。图9-19b所示结果表明，与经典的体硅工艺相比，使用 FinFET 和 UTB 器件的 FO4 反相器的功耗得到明显减小。使用 FinFET 减小的能量损耗可以高达60%。由于 FinFET 结构具有优良的可缩放性和重要的电路性能优势，在 65nm 技术节点（约 25nm 的物理栅长）的 IC 制造中就会采用[102]。

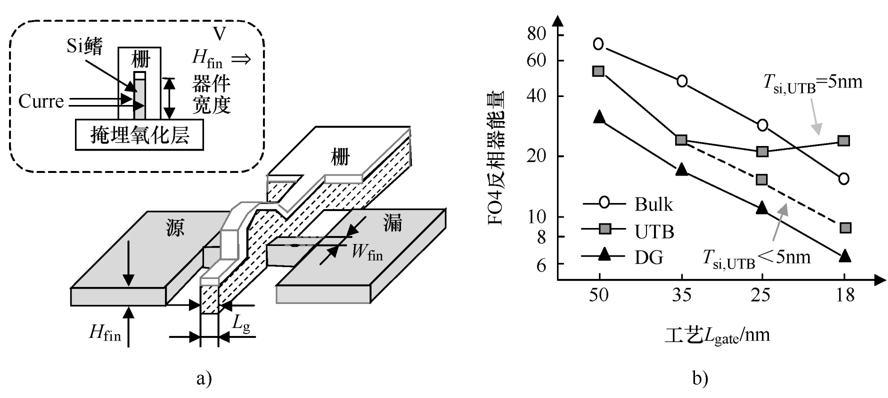


图 9-19 FinFET 结构和潜在的节省有源功率的优点

a）FinFET 三维示意图 b）几种技术之间的性能比较（来源于参考文献［102］）

除了 SOI 和双栅器件，其他的技术革新包括低 k 介质（气隙 Cu 技术）和为 90nm 以及更小工艺新开发的有氮化硅“帽”的应变硅技术。

9.4.2 组装技术革新

由于本身存在较大的电容，所以 I/O 系统消耗了大量的功率。封装衬底和印制电路板中的互连比芯片上的互互连长得多，与它们相关的电容达到芯片上电容的两倍。新型的组装技术和接口信号设计技术，例如系统级封装（SiP）、三维集成、RF 无线连接和光学互连[98]，给低功耗系统集成提供了希望。在这些开发的技术中，SiP 方法可以将多个不同种类的芯片和 RF 无源元件集成在同一个封装结构中[99]，极大地减小了 I/O 功耗而实际上不要求开发新的设计原理。

9.5　未来超低功耗设计的展望

随着具有优越便携性能和智能化特点的计算/通信设备的不断广泛使用和持续发展，这一趋势将继续推动系统设计进入更低功耗的阶段。除了前面介绍的技术进步可以进一步适应器件尺寸的缩小以满足下一代应用的需要，本节介绍对未来超低功耗设计影响最重要的几种可能领域和技术。

9.5.1　亚阈区电路工作

由于亚阈区电路工作的功耗远小于通常的强反型电路，对未来采用非电源供电的设计来说是一个强有力的待选对象。与常规的 CMOS 逻辑相比，亚阈区电路还具有跨导增益大以及近似理想的静态噪声容限的优势。然而，由于没有反型导电沟道，如果不能采取适当的控制措施，亚阈区电路对电源电压、温度和工艺波动的敏感度将高得令人却步，这就限制了这一技术的近期使用。在克服这些困难的努力中，人们提出了一些亚阈区逻辑，包括可变 V_{th}亚阈区 CMOS（VT-sub-CMOS）和亚阈区动态 V_{th} MOS（sub-DTMOS）逻辑。如图 9-20 所示，VT-sub-CMOS 逻辑采用了一个附加的稳定性设计方案，用一个稳定电路来监测因温度和工艺波动引起的晶体管电流的任何变化，并且给衬底施加一个合适的偏置。VT-sub-CMOS 的逻辑电路和稳定电路都工作在亚阈区。9.3.2 节中引入的 DTMOS 逻辑对低电压工作来说是一个引人注目的待选者。与亚阈区 CMOS 逻辑相比，sub-DTMOS 的栅电容较大，但是它提供有高得多的动态电流。这两个亚阈区逻辑系列的功率延迟积（PDP）相似，但是在维持相同的能量/翻转比的情况下，sub-DTMOS 可以工作在较高的开关频率下。VT-sub-CMOS 和 sub-DTMOS 对工艺和温度的波动都可以实现期望的鲁棒性和容限，但是付出的代价是要采用额外的稳定电路，并增加了工艺的复杂度[32]。

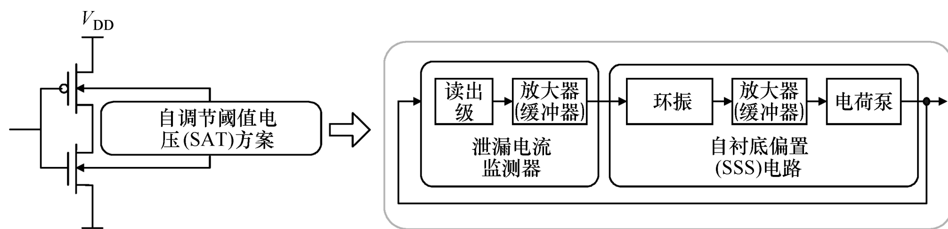


图 9-20　采用了稳定方案的 VT-sub-CMOS 逻辑（来源于参考文献［32］）

9.5.2 容错设计

随着器件尺寸的缩小和电源电压的降低，未来的器件和互连将存在会遭遇更大的工艺波动以及更容易遭受到如自然辐射和电噪声等外部干扰攻击的弱点。将工作中要求 100% 的正确率放宽到允许适当的错误率，可以极大的减小设计费用，但是同时要求具有一定容错能力的可靠的超低功耗设计。迄今为止，不同设计级别和应用领域中已采用了许多容错方案，例如，DRAM 设计和通信过程中的纠错码（Error Correction Code，ECC）、采用硬件（三重表决方案（TMR）、看门狗处理器设计、和动态执行验证结构（Dynamic Implementation Verification Architecture，DIVA）[33]）和软件方式（同步冗余线程处理器（SRT）[34]）冗余方法的计算机结构验证方案等。未来的鲁棒性超低功耗系统将如目前的系统低功耗设计方式那样，将综合使用各种容错方案。

9.5.3 异步设计与同步设计

过去的 20 年中，同步时序方法已经成功实现了 VLSI 设计的规模按照指数规律持续增长，并取得了现代处理器设计、构建完善的设计方法学、以及解决和优化同步问题的先进计算机辅助设计工具等方面的杰出成就。然而，当设计者的目标进一步迈向几 GHz 的高工作频率和复杂度增加的更大系统，而同时允许的功耗却受到制约，常规的同步方法不可避免地遇到了严重的问题。时钟控制不确定性和时钟分布网络的功耗是减小系统设计费用的主要阻碍。工艺波动的增加也损害着同步系统的性能，更严重的是在所有的其他环境中必须执行最坏情况时序。由于在提高功率的利用效率方面异步设计具有相当的潜力，最近几年异步设计方法又重新引起人们注意，得到了广泛的研究[80-83]。与同步方法相比，异步设计的特点是仅在有效工作的情况消耗功率、针对典型情况而不是最坏情况优化子组元件、更低的噪声和电磁辐射、以及全局时序协调的困难较小[80,81]。异步设计特别适合于计算负载波动具有不确定性以及最坏情况与典型条件下性能有巨大差别的应用[80]。目前从完全的异步电路到全局异步和局部同步（Globally Asynchronous and Locally Synchronous，GALS）[84] 系统，都存在着可能的解决方法。GALS 源于同步结构的改进。作为一个中间级，GALS 减小了设计方法转换的困难，占用更小的面积，但是要比全异步方式消耗更多的功率[82]。

9.5.4 栅感应泄漏抑制方法

在 130nm 技术中，栅泄漏电流仍然不是主要的泄漏部分。但是随着薄氧化层厚度 t_{OX} 不断快速缩小的趋势，栅氧隧穿泄漏和栅感应漏泄漏（GIDL）很快就会达到与亚阈区电流相比拟的程度。为了将其置于可控状态，需要采用有效的技

术。传统的泄漏减小技术，例如减小电源电压和关断未使用的部分，仍然是应对新泄漏成分的有效手段。栅泄漏抑制的现有其他方法包括引脚重排[36]和电场弛豫（Electrical Field Relaxation，EFR）[37]方法。引脚重排技术利用了栅泄漏与非传导层叠中“关断”器件位置的依赖关系。引脚重排优化结果表明，待机栅泄漏得到22%～82%的减小，并且运行器件的栅泄漏也得到高达25%的减小[36]。通过使 SRAM 单元晶体管的栅漏电压从 1.5V 减小到 1V[37]，EFR 方法使 GIDL 电流减小了 90%。应用双 t_{OX}是在未来高速、低功耗 DRAM 设计中建议采用的一项技术，其中外围电路中采用薄 t_{OX}有助于实现更快的操作，而核心单元采用厚 t_{OX}确保了稳定的工作并抑制栅隧穿泄漏电流。同样，采用双 V_{th}和双 V_{DD}可以满足 RAM 单元和外围电路的不同要求，实现对存储器性能和功耗的优化设计。除了电路改善，未来在技术层次的创新，例如新的有低泄漏和高介质常数特点的栅介质材料的开发，可能是最期望的方法[103]。

参 考 文 献

[1] V. De and S. Borkar, Technology and design challenges for low power and high performance, *International Symposium on Low Power Electronics and Design,* pp. 163–168, Aug. 1999.

[2] T. Sakurai, Perspectives on power-aware electronics, *IEEE International Solid-State Circuits Conference*, pp. 1–16, Feb. 2003.

[3] B. Chatterjee et al., Effectiveness and scaling trends of leakage control techniques for sub-130 nm CMOS technologies, *International Symposium on Low Power Electronics and Design,* pp. 122–127, Aug. 2003.

[4] P. J. M. Havinga and G. J. M. Smit, Design techniques for low power systems, *J. Syst. Archit.*, Vol. 46, No. 1, 2000.

[5] M. Sheets et al., Power management for PicoRadio, *Gigascale Systems Research Center Workshop*, June 2002.

[6] K. Itoh, K. Sasaki, and Y. Nakagome, Trends in low-power RAM circuit technologies, *Proc. IEEE*, pp. 524–543, Apr. 1995.

[7] J. P Fishburn and S. Taneja, Transistor sizing for high performance and low power, *Custom Integrated Circuits Conference*, pp. 591–594, May, 1997.

[8] M. Hamada, Y. Ootaguro, and T. Kuroda, Utilizing surplus timing for power reduction, *Custom Integrated Circuits Conference*, pp. 89–92, May, 2001.

[9] R. Brodersen et al., Methods for true power minimization, *International Conference on Computer Aided Design*, Nov. 2002.

[10] C. Piguet et al., Low-power low-voltage library cells and memories, *IEEE International Conference on Electronics, Circuits and Systems*, Vol. 3, pp. 1521–1524, Sept. 2001.

[11] H. Mizuno and T. Nagano, Driving source-line (DSL) cell architecture for sub-1-V high-speed low-power applications, *Digest of Technical Papers, Symposium on VLSI Circuits*, pp. 25–26, June 1995.

[12] K. Itoh, A. R. Fridi, A. Bellaouar, and M. I. Elmasry, A deep sub-V, single power-supply, SRAM cell with multi-Vt, boosted storage node and dynamic load, *Digest of Technical Papers, Symposium on VLSI Circuits,* pp. 132–133, June 1996.

[13] H. Yamauchi, T. Iwata, H. Akamatsu, and A. Matsuzawa, A 0.5 V single power supply operated high-speed boosted and offset-grounded data storage (BOGS) SRAM cell architecture, *IEEE Trans. VLSI Syst.*, Vol. 5, No. 4, pp. 377–387, Dec. 1997.

[14] O. Minato et al., A 20 ns 64 K CMOS RAM, *IEEE International Solid-State Circuits Conference*, pp. 222–223, Feb. 1984.

[15] J. S. Caravella, A low voltage SRAM for embedded applications, *IEEE J. Solid-State Circuits*, Vol. 32, No. 3, pp. 428–432, Mar. 1997.

[16] M. Yoshimoto et al., A 64 Kb full CMOS RAM with divided word line structure, *IEEE International Solid-State Circuits Conference*, Vol. XXVI, pp. 58–59, Feb. 1983.

[17] B. S. Amrutur and M. A. Horowitz, Techniques to reduce power in fast wide memories, *Proc. SLPE'94*, pp. 92–93, 1994.

[18] T. Mori et al., A 1 V 0.9 mW at 100 MHz 2 k × 16 b SRAM utilizing a half-swing pulsed-decoder and write-bus architecture in 0.25 μm dual-Vt CMOS, *IEEE International Solid-State Circuits Conference*, pp. 22.4-1–22.4-2, Feb. 1998.

[19] K. W. Mai et al., Low-power SRAM design using half-swing pulse-mode techniques, *IEEE J. Solid-State Circuits*, Vol. 33, No. 11, pp. 1659–1671, Nov. 1998.

[20] S. Flannagan et al., Two 64 K CMOS SRAMs with 13 ns access time, *IEEE International Solid-State Circuits Conference*, Vol. XXIX, pp. 208–209, Feb. 1986.

[21] H. Aydin et al., Dynamic and aggressive scheduling techniques for power-aware real-time systems, *Real-Time Systems Symposium*, London, Dec. 2001.

[22] L. Benini, A. Bogliolo, and G. De Micheli, A survey of design techniques for system-level dynamic power management, *IEEE Trans. VLSI Syst.*, Vol. 8, No. 3, pp. 299–316, June 2000.

[23] J. Gomez, A. T. Campbell, M. Naghshineh, and C. Bisdikian, Power-aware routing in wireless packet networks, *IEEE International Workshop on Mobile Multimedia Communications*, pp. 380–383, Nov. 1999.

[24] T. D. Burd, T. A. Pering, A. J. Stratakos, and R. W. Brodersen, A dynamic voltage scaled microprocessor system, *IEEE J. Solid-State Circuits*, Vol. 35, No. 11, pp. 1571–1580, Nov. 2000.

[25] C. H. Kim and K. Roy, Dynamic V_{TH} scaling scheme for active leakage power reduction design, *Proceedings of Design, Automation and Test in Europe Conference and Exhibition*, pp. 163–167, Mar. 2002.

[26] S. Lee and T. Sakurai, Run-time voltage hopping for low-power real-time systems, *Design Automation Conference*, pp. 806–809, June 2000.

[27] S. Narendra et al., Scaling of stack effect and its application for leakage reduction, *International Symposium on Low Power Electronics and Design*, pp. 195–200, Aug. 2001.

[28] S. Kaxiras, Z. Hu, and M. Martonosi, Cache decay: exploiting generational behavior to reduce cache leakage power, *International Symposium on Computer Architecture*, pp. 240–251, June–July 2001.

[29] K. Flautner et al., Drowsy caches: simple techniques for reducing leakage power, *International Symposium on Computer Architecture*, pp. 148–157, May 2002.

[30] H. Qin et al., SRAM leakage suppression by minimizing standby supply voltage, *IEEE International Symposium on Quality Electronic Design*, Mar. 2004.

[31] M. Margala, Low-power SRAM circuit design, *IEEE International Workshop on Memory Technology, Design and Testing*, pp. 115–122, Aug. 1999.

[32] H. Soeleman, K. Roy, and B. C. Paul, Robust subthreshold logic for ultra-low power operation, *IEEE Trans. VLSI Syst.*, Vol. 9, No. 1, pp. 90–99, Feb. 2001.

[33] T. M. Austin, DIVA: a reliable substrate for deep submicron microarchitecture design, *ACM/IEEE International Symposium on Microarchitecture*, 1999.

[34] S. K. Reinhardt and S. S. Mukherjeem, Transient fault detection via simultaneous multithreading, *International Symposium on Computer Architecture*, 2000.

[35] A. Ono et al., A 100 nm node CMOS technology for practical SOC application requirement, *IEEE International Electron Devices Meeting*, pp. 511–514, 2001.

[36] D. Lee, W. Kwong, D. Blaauw, and D. Sylvester, Analysis and minimization techniques for total leakage considering gate oxide leakage, *Design Automation Conference*, pp. 175–180, June 2003.

[37] K. Osada, Y. Saitoh, E. Ibe, and K. Ishibashi, 16.7fA/cell tunnel-leakage-suppressed 16-Mbit SRAM based on electric-field-relaxed scheme and alternate ECC for handling cosmic-ray-induced multi-errors, *IEEE International Solid-State Circuits Conference*, pp. 260–261, Feb. 1996.

[38] J. P. Fishburn and A. E. Dunlop, TILOS: a posynomial programming approach to transistor sizing, *International Conference on Computer-Aided Design*, pp. 326–328, Nov. 1985.

[39] A. R. Conn et al., Gradient-based optimization of custom circuits using a static-timing formulation, *Design Automation Conference*, pp. 452–459, June 1999.

[40] M. Borah, R. Owens, and M. Irwin, Transistor sizing for low power CMOS circuits, *IEEE Trans. Comput. Aided Des. Integrated Circuits Syst.*, Vol. 15, No. 6, 665–671, 1996.

[41] S. Ma and P. Franzon, Energy control and accurate delay estimation in the design of CMOS buffers, *IEEE J. Solid-State Circuits,* Vol. 29, No. 9, pp. 1150–1153, Sept. 1994.

[42] J. P. Halter and F. N. Najm, A gate-level leakage power reduction method for ultra-low-power CMOS circuits, *IEEE Custom Integrated Circuits Conference*, pp. 475–478, May 1997.

[43] Z. Chen, M. Johnson, L. Wei, and W. Roy, Estimation of standby leakage power in CMOS circuit considering accurate modeling of transistor stacks, *International Symposium on Low Power Electronics and Design*, pp. 239–244, Aug. 1998.

[44] M. C. Johnson, D. Somasekhar, L. Chiou, and K. Roy, Leakage control with efficient use of transistor stacks in single threshold CMOS, *IEEE Trans. VLSI Syst.*, Vol. 10, No. 1, pp. 1–5, Feb. 2002.

[45] A. Abdollahi, F. Fallah, and M. Pedram, Runtime mechanisms for leakage current reduction in CMOS VLSI circuits, *International Symposium on Low Power Electronics and Design*, pp. 213–218, Aug. 2002.

[46] Y. Ye, S. Borkar, and V. De, A new technique for standby leakage reduction in high-performance circuits, *Digest of Technical Papers, Symposium on VLSI Circuits*, pp. 40–41, 1998.

[47] E. Musoll and J. Cortadella, Optimizing CMOS circuits for low power using transistor reordering, *European Design and Test Conference*, pp. 219–223, Mar. 1996.

[48] S. C. Prasad and K. Roy, Circuit optimization for minimization of power consumption under delay constraint, *International Conference on VLSI Design*, pp. 305–309, Jan. 1995.

[49] R. Hossain, M. Zheng, and A. Albicki, Reducing power dissipation in CMOS circuits by signal probability based transistor reordering, *IEEE Trans. Comput. Aided Des. Integrated Circuits Syst.*, Vol. 15, No. 3, pp. 361–368, Mar. 1996.

[50] W. Z. Shen, J. Y. Lin, and F. W. Wang, Transistor reordering rules for power reduction in CMOS gates, *Asian and South Pacific Design Automation Conference*, pp. 1–6, Aug. 1995.

[51] M. Hashimoto, H. Onodera, and K. Tamaru, Input reordering for power and delay optimization, *IEEE International ASIC Conference and Exhibit*, pp. 194–199, Sept. 1997.

[52] H. Onodera, M. Hashimoto, and T. Hashimoto, ASIC design methodology with on-demand library generation, *Digest of Technical Papers, Symposium on VLSI Circuits*, pp. 57–60, June 2001.

[53] M. Hashimoto and H. Onodera, Post-layout transistor sizing for power reduction in cell-based design, *Asia and South Pacific Design Automation Conference*, pp. 359–365, Feb. 2001.

[54] S. Sirichotiyakul et al., Duet: an accurate leakage estimation and optimization tool for dual-V_t circuits, *IEEE Trans. VLSI Syst.*, Vol. 10, No. 2, pp. 79–90, Apr. 2002.

[55] Z. Chen et al., 0.18 μm dual V_t MOSFET process and energy-delay measurement, *International Electron Devices Meeting*, pp. 851–854, Dec. 1996.

[56] K. Fujii, T. Douseki, and M. Harada, A sub-1 V triple-threshold CMOS/SIMOX circuit for active power reduction, *IEEE International Solid-State Circuits Conference*, pp. 190–191, Feb. 1998.

[57] L. Wei et al., Design and optimization of low voltage high performance dual threshold CMOS circuits, *Design Automation Conference*, pp. 489–494, June 1998.

[58] K. Nose and T. Sakurai, Optimization of V_{DD} and V_{TH} for low-power and high-speed applications, *Asia and South Pacific Design Automation Conference*, pp. 469–474, Jan. 2000.

[59] M. Ukita et al., A single-bit-line cross-point cell activation (SCPA) architecture for ultra-low-power SRAM's, *IEEE J. Solid-State Circuits*, Vol. 28, No. 11, pp. 1114–1118, Nov. 1993.

[60] S. Manne, A. Klauser, and D. Grunwald, Pipeline gating: speculation control for energy reduction, *International Symposium on Computer Architecture*, pp. 132–141, July 1998.

[61] O. Minato et al., A 20 ns 64 K CMOS static RAM, *IEEE J. Solid-State Circuits*, Vol. 19, No. 6, pp. 1008–1013, Dec. 1984.

[62] K. Ishibashi et al., A 9-ns 1-Mbit CMOS SRAM, *IEEE J. Solid-State Circuits*, Vol. 24, No. 5, pp. 1219–1225, Oct. 1989.

[63] E. Seevinck, A current sense-amplifier for fast CMOS SRAMs VLSI circuits, *Digest of Technical Papers, Symposium on VLSI Circuits*, pp. 71–72, June 1990.

[64] K. Sasaki et al., 7-ns 140-mW 1-Mb CMOS SRAM with current sense amplifier, *IEEE J. Solid-State Circuits*, Vol. 27, No. 11, pp. 1511–1518, Nov. 1992.

[65] S. Douseki et al., 1-V power supply high-speed digital circuit technology with multithreshold-voltage CMOS, *IEEE J. Solid-State Circuits*, Vol. 30, No. 8, pp. 847–854, Aug. 1995.

[66] T. Kuroda et al., A 0.9-V, 150-MHz, 10-mW, 4 mm^2, 2-D discrete cosine transform core processor with variable threshold-voltage (VT) scheme, *IEEE J. Solid-State Circuits*, Vol. 31, No. 11, pp. 1770–1779, Nov. 1996.

[67] J. Kao, S. Narendra, and A. Chandrakasan, MTCMOS hierarchical sizing based on mutual exclusive discharge patterns, *Design Automation Conference*, pp. 495–500, June 1998.

[68] H. Kawaguchi, K. Nose, and T. Sakurai, A super cut-off CMOS (SCCMOS) scheme for 0.5-V supply voltage with picoampere stand-by current, *IEEE J. Solid-State Circuits*, Vol. 35, No. 10, pp. 1498–1501, Oct. 2000.

[69] T. Inukai et al., Boosted gate MOS (BGMOS): device/circuit cooperation scheme to achieve leakage-free giga-scale integration, *IEEE Custom Integrated Circuits Conference*, pp. 409–412, May 2000.

[70] T. Kuroda, Low power CMOS digital design for multimedia processors, *International Conference on VLSI and CAD*, pp. 359–367, Oct. 1999.

[71] K. Kanda, K. Nose, H. Kawaguchi, and T. Sakurai, Design impact of positive temperature dependence on drain current in sub-1-V CMOS, *IEEE J. Solid-State Circuits*, Vol. 36, No. 10, pp. 1559–1564, Oct. 2001.

[72] F. Assaderaghi et al., A dynamic threshold voltage MOSFET (DTMOS) for ultra-low voltage operation, *International Electron Devices Meeting*, pp. 809–812, Dec. 1994.

[73] F. Assaderaghi, DTMOS: its derivatives and variations, and their potential applications in microelectronics, *International Conference on Microelectronics*, pp. 9–10, Oct. 2000.

[74] D. M. Brooks et al., Power-aware microarchitecture: design and modeling challenges for next-generation microprocessors, *IEEE Micro Mag.*, Vol. 20, No. 6, pp. 26–44, Nov. 2000.

[75] G. E. Tellez, A. Farrahi, and M. Sarrafzadeh, Activity-driven clock design for low power circuits, *IEEE/ACM International Conference on Computer-Aided Design*, pp. 62–65, Nov. 1995.

[76] K. S. Min, H. Kawaguchi, and T. Sakurai, Zigzag super cut-off CMOS (ZSCCMOS) block activation with self-adaptive voltage level controller: an alternative to clock-gating scheme in leakage dominant era, *IEEE International Solid-State Circuits Conference*, pp. 1–10, Feb. 2003.

[77] M. Horiguchi, T. Sakata, and K. Itoh, Switched-source-impedance CMOS circuit for low standby subthreshold current giga-scale LSI's, *IEEE J. Solid-State Circuits*, Vol. 28, No. 11, pp. 1131–1135, Nov. 1993.

[78] H. Kawaguchi et al., Dynamic leakage cut-off scheme for low-voltage SRAMs, *Digest of Technical Papers, Symposium on VLSI Circuits*, pp. 140–141, June 1998.

[79] K. Nii et al., A low power SRAM using auto-backgate-controlled MT-CMOS, *International Symposium on Low Power Electronics and Design*, pp. 293–298, Aug. 1998.

[80] N. C. Paver and D. A. Edwards, Is asynchronous logic good for low-power? *IEE Colloquium on Low Power Analogue and Digital VLSI: ASICS, Techniques and Applications*, pp. 4/1–4/5, June 1995.

[81] C. H. Van Berkel, M. B. Josephs, and S. M. Nowick, Applications of asynchronous circuits, *Proc. IEEE*, Vol. 87, No. 2, pp. 223–233, Feb. 1999.

[82] C. Piguet, M. Renaudin, and T. J.-F. Omnes, Special session on low-power systems on chips (SOCs), *Conference and Exhibition on Design, Automation and Test in Europe*, pp. 488–494, Mar. 2001.

[83] V. G. Oklobdzija and J. Sparso, Future directions in clocking multi-GHz systems, *International Symposium on Low Power Electronics and Design*, p. 219, Aug. 2002.

[84] D. M. Chapiro, Globally-asynchronous locally-synchronous systems, Ph.D dissertation, Stanford University, Oct. 1984.

[85] M. Miyazaki et al., A 1000-MIPS/W microprocessor using speed-adaptive threshold-voltage CMOS with forward bias, *IEEE International Solid-State Circuits Conference*, pp. 420–421, Feb, 2000.

[86] J. Tschanz, Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage, *IEEE International Solid-State Circuits Conference*, pp. 422–423, Feb. 2002.

[87] T. Sugibayashi et al., A 30 ns 256 Mb DRAM with multi-divided array structure, *IEEE International Solid-State Circuits Conference*, pp. 24–26, Feb. 1993.

[88] K. Kennnizaki et al., A 36/spl mu/A 4 Mb PSRAM with quadruple array operation, *Digest of Technical Papers, Symposium on VLSI Circuits*, pp. 79–80, May 1989.

[89] N. C.-C. Lu and H. H. Chao, Half-V/SUB DD/bit-line sensing scheme in CMOS DRAMs, *IEEE International Solid-State Circuits Conference*, Vol. 19, No. 4, pp. 451–454, Aug. 1984.

[90] T. Kawahara et al., A charge recycle refresh for Gb-scale DRAM's in file applications, *IEEE International Solid-State Circuits Conference*, Vol. 29, No. 6, pp. 715–722, June 1994.

[91] K. Itoh, Low-voltage memories for power-aware systems, *International Symposium on Low Power Electronics and Design*, pp. 1–6, Aug. 2002.

[92] K. Itoh, *VLSI Memory Chip Design*, Springer-Verlag, New York, 2001.

[93] D. W. Bailey and B. J. Benschneider, Clocking design and analysis for a 600-MHz alpha microprocessor, *IEEE J. Solid-State Circuits*, Vol. 33, pp. 1627–1633, Nov. 1998.

[94] J. Wood, T. C. Edwards, and S. Lipa, Rotary traveling-wave oscillator arrays: a new clock technology, *IEEE J. Solid-State Circuits*, Vol. 36, pp. 1654–1665, Nov. 2001.

[95] F. O'Mahony, C. P. Yue, M. Horowitz, and S. S. Wong, 10 GHz clock distribution using coupled standing-wave oscillators, *International Solid-State Circuits Conference*, pp. 1–4, 2003.

[96] S. C. Chan, K. L. Shepard, and P. J. Restle, Design of resonant global clock distributions, *International Conference on Computer Design*, pp. 248–253, 2003.

[97] M. Igarashi et al., A diagonal-interconnect architecture and its application to RISC core design, *International Solid-State Circuits Conference*, pp. 272–273, 2002.

[98] J. D. Meindl et al., Interconnecting device opportunities for gigascale integration (GSI), *IEEE International Electron Devices Meeting*, pp. 525–528, 2001.

[99] K. L. Tai, System-in-package (SIP): challenges and opportunities, *IEEE Asia and South Pacific Design Automation Conference*, pp. 191–196, Jan. 2000.

[100] Y.-K. Choi, K. Asano, N. Lindert, V. Subramanian, T.-J. King, J. Bokor, and C. Hu, Ultra-thin body SOI MOSFET for deep-subtenth micron era, *Technical Digest, IEEE International Electron Devices Meeting*, pp. 919–921, 1999.

[101] X. Huang, W.-C. Lee, C. Kuo, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Y.-K. Choi, K. Asano, V. Subramanian, T.-J. King, J. Bokor, and C. Hu, Sub-50 nm FinFET: PMOS, *Technical Digest, IEEE International Electron Devices Meeting*, pp. 67–70, 1999.

[102] L. Chang, Y.-K. Choi, D. Ha, P. Ranade, S. Xiong, J. Bokor, C. Hu, and T.-J. King, Extremely scaled silicon nano-CMOS devices, *Proc. IEEE*, Vol. 91, No. 11, pp. 1860–1873, Nov. 2003.

[103] Y. Nakagome, M. Horiguchi, T. Kawahara, and K. Itoh, Review and future prospects of low-voltage RAM circuits, *IBM J. Res. Dev.*, Vol. 47, No. 5/6, 2003.

[104] B. Wong, Method to reduce leakage during a semiconductor burn-in procedure, U.S. patent 6,649,425, Nov. 18, 2003.

第 10 章　可制造性设计

10.1　引言

随着纳米 CMOS 技术特征尺寸的缩小，工艺可制造性已经跟不上缩小的要求。进入亚波长阶段，这种差距变得更大，使得它很难实现大多数图形结构的光刻[4]。一些图形结构甚至更难实现完美光刻，这会导致光学失真，而这些失真在某些情况下会引起成品率下降和性能退化[2]。业界迄今依然依赖于光学邻近修正（OPC）和其他分辨率扩展技术（RET）来解决亚波长阶段的这种差距（见第 3 章中 OPC 和 RET 的内容）。然而，OPC 的修正能力是有限的，这就将芯片开发过程中提高设计成品率这一重要任务留给了设计者。设计者必须充分了解光刻步骤，以生成失真最少的版图，并且将这些知识应用到设计中。

互连制造问题是纳米 CMOS 工艺中对成品率影响最大的因素。一个没有考虑可制造性的设计会导致铜的腐蚀和凹陷，从而改变影响电迁移和时序的设计特性。这将使速度下降并且因此导致产品的平均售价降低。有些互连图形会因短路而导致成品率下降。通孔开路是铜工艺中的另一种成品率破坏因素。互连密度的变化引起了层间介质（ILD）厚度的变化（请参见图 11-15），由于抛光不足导致的金属短路以及容性寄生变化引起的不期望的时序[2]，将导致成品率损失。

多晶特征尺寸（poly-CD）受多晶密度和节距的影响，会导致无法预知的时序[1,3]。由较长 poly-CD 引起的传输时延超过最大允许时间会导致速度下降。在最小时延路径上，窄 poly-CD 引起的最小时延时间会导致元器件失去功能，在设计中必须认真对待。因为这种 poly-CD 的差异是系统级的，在一定程度上可以在光刻中采用光学方法修正（详细情况见第 3 章）。这个修正将要求进行图形密度分析和掩膜的变化。如果设计中晶体管都按照同一方向排列，这种修正会相对容易。否则，制造工程师可能需要重复进行多次试验，以确定对两个方向晶体管的最优修正。这种制造过程中的修正可能会推迟产品推出的时间。

由于栅损伤，天线效应可能会导致成品率损失，并且某种情况下，由于更早地出现负偏压温度不稳定性（NBTI）引起的 V_{th}漂移，会使晶体管性能衰退。因

此，纳米 CMOS 工艺需要改变设计方法以提高电路制造的成品率，同时在设计中尽量少地引入分散性和寄生参数。下面三节中我们将通过分析几个设计案例阐明这种需求。

10.2　最优和亚最优版图对比

下面的案例研究分别说明好的和差的版图实践。在第 11 章中我们将探究版图对器件参数分散性的影响以及如何避免这些问题。

图 10-1a 所示为是一个库中门电路的两种不同版图。位于版图中的不同的区域使这个门变得不可缩小并且增加了它对工艺变化和晶体管末端泄漏的敏感性。最小扩散窗口的工艺变形导致端帽覆盖的损失。这又将导致严重的晶体管泄漏。

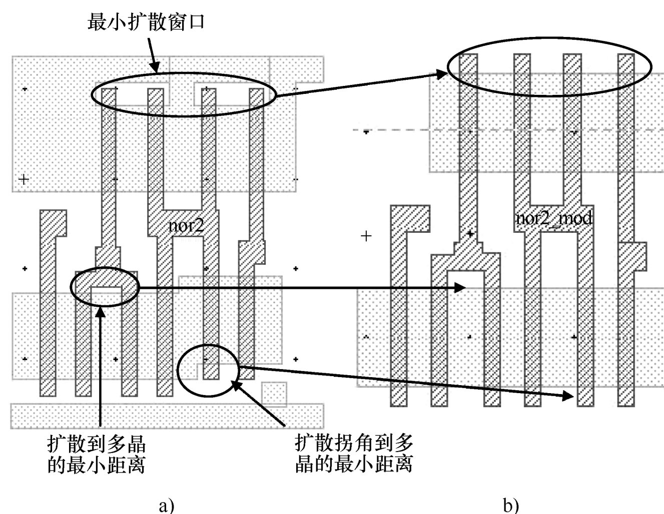


图 10-1　a）一个库中的门电路的版图　b）改进的版图

与反 U 型多晶硅之间的最小扩散间距导致了晶体管平均长度更长，并且因此使驱动电流更小。在理想对准情况下，扩散拐角到多晶硅的最小间距刚刚能使端帽覆盖，但是如果不能完美对准，出现偏移，即使这种情况下扩散拐角是小的急弯，也不能保证端帽覆盖。如果这个急弯的内拐角较长，情况将更加严重。这个单元是为 130nm 节点设计的。缩小到 90nm 时，光刻的单元最终虚像如图 10-2 所示。同一单元的改进版图如图 10-1b 所示，其中对原来单元中所有很差的且会引起失效的区域进行了改进。

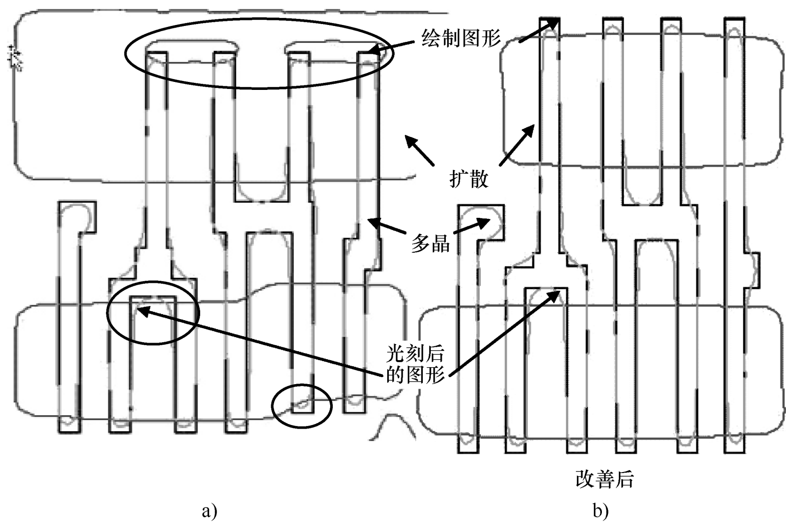


图 10-2 a）图 10-1 中单元的光刻最终虚像 b）改进版图的效果

接触孔和通孔开口是另一种影响成品率的因素。因此应该谨慎使用两个中间有间距的接触孔。图 10-3 显示的是一个版图实例以及改进方法。图 10-4 是使用单个通孔的单元；重画了金属 1 与通孔相连的接触区，使其大于通孔面积，将能够改善产品的成品率。

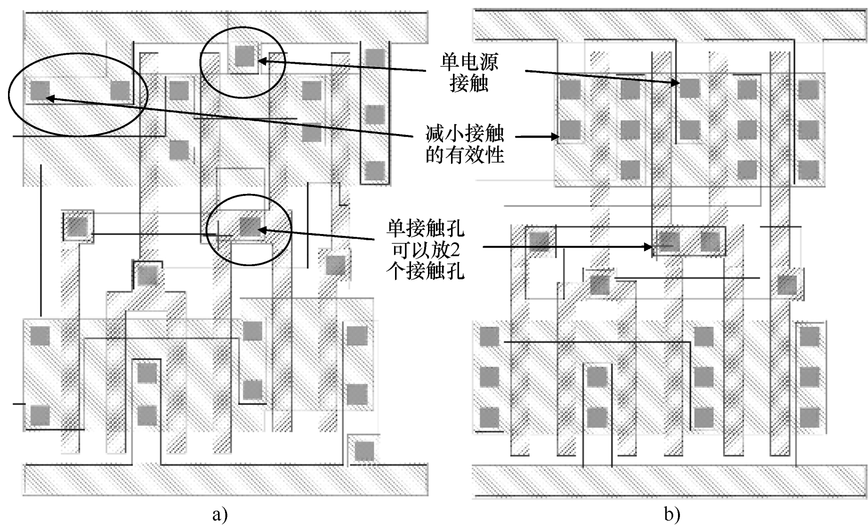


图 10-3 a）采用两个接触孔的版图 b）改进的版图

对纳米 CMOS 技术，如果设计中所有晶体管都朝一个方向的话，控制 poly-CD 应该更容易。如果需要偏置来修正光学和刻蚀变形，只要使晶体管多晶同向

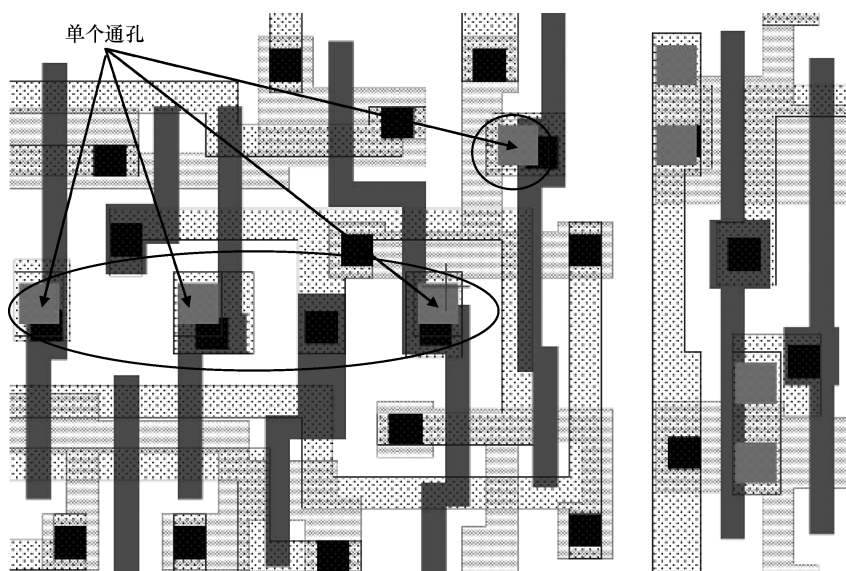


图 10-4　采用单个通孔的版图

排列，在设计中就能够比较容易地实现上述修正要求。这点对模拟电路、存储器位单元、读出放大器和其他关键电路来说特别重要。图 10-5a 是晶体管垂直排向和水平排向的版图。图 10-5b 是将所有晶体管同一排向的版图。这个版图还包括了前面我们讨论的其他改善。

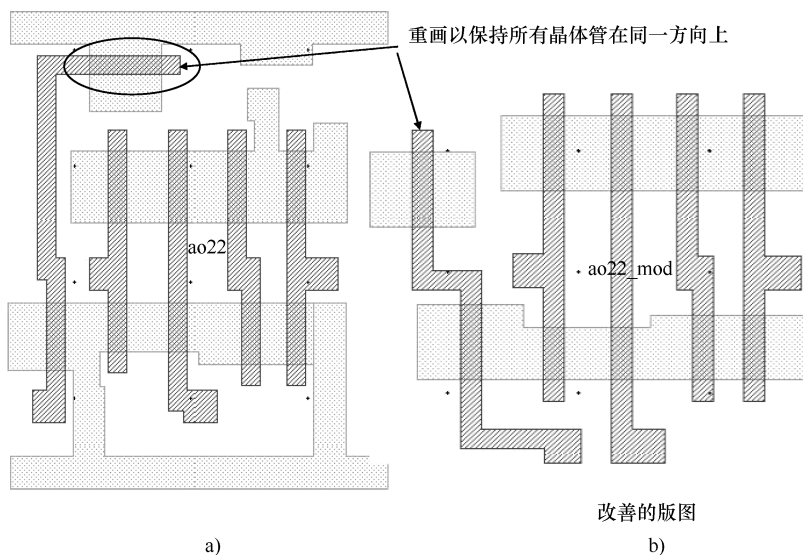


图 10-5　a）包括垂直和水平方向晶体管的版图　b）改进的版图

图 10-6 所示为由于未对准和扩散图形光刻后直角弧化导致短路的版图。图中节点 X 处是扩散图形光刻后直角弧化，如果存在未对准情况使得多晶 Y 与节点 X 处的扩散交迭，就会引起节点 A 和节点 B 短路。这个版图上有两个地方会发生这种情况，如图 10-6 中圆圈所示的范围。

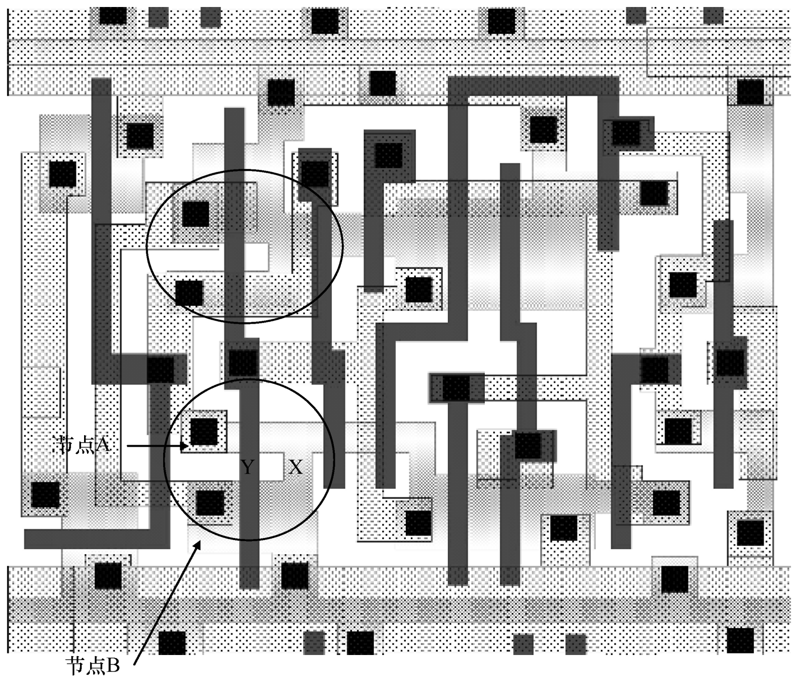


图 10-6　导致短路的版图

图 10-7a 所示为一个设计得很差的触发器的光刻仿真虚像。这个触发器的版图如图 10-8 所示。图 10-8 中所示的采用最小扩散间距的 C 型扩散会因为扩散图形光刻后直角弧化而发生扩散短路。这个短路在图 10-7a 所示的虚像中非常明显。改进后没有 C 型扩散区的版图则不会引起短路（见图 10-7b）。

图 10-9 所示为另一种对光刻工程师来说具有挑战性的版图。最小的 U 型扩散槽不仅很难被光刻出来，而且也很难被缩小到下一个工艺节点中。由于扩散图形光刻后直角弧化，像触发器例子中的 C 型扩散区那样（见图 10-8），它也易于形成短路。

图 10-10 所示为一个接触孔太靠近扩散区边界的情况，这里扩散与 T 型多晶之间采用的是最小间距。多晶图形光刻后直角弧化加之未对准就会导致多晶与接触孔短路。这些问题的解决方法是增大多晶到扩散边界的间距，从而多晶图形光刻后直角弧化不会扩展到扩散上。

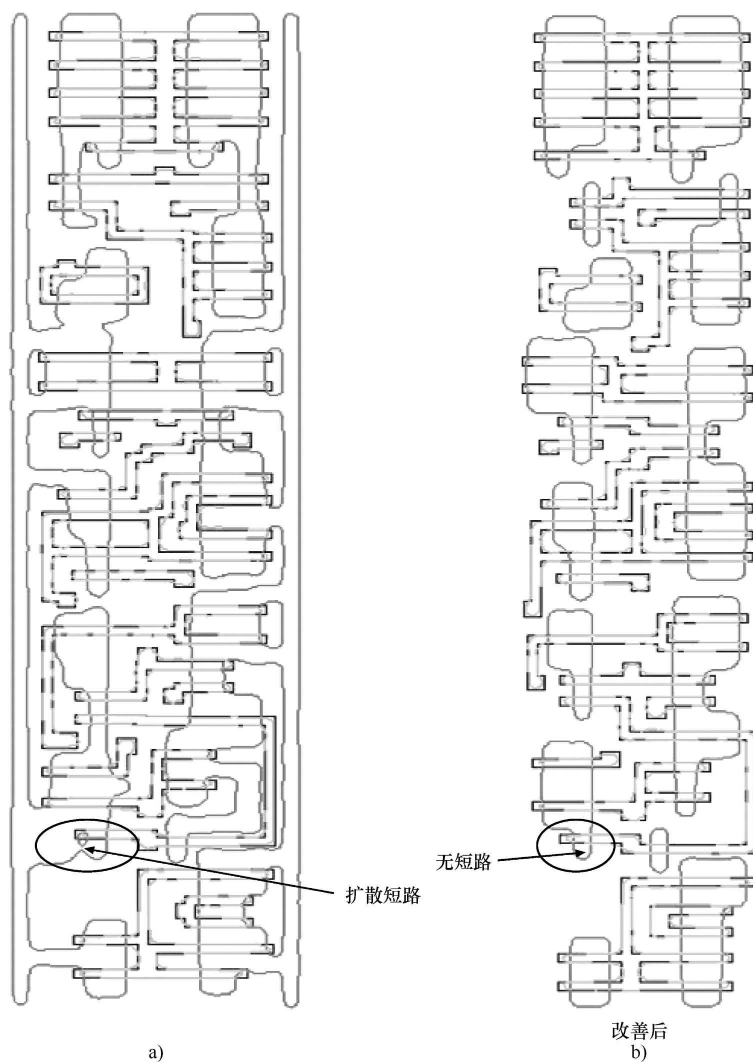


图 10-7　a）一个设计得很差的触发器的光刻仿真虚像　b）改进的版图

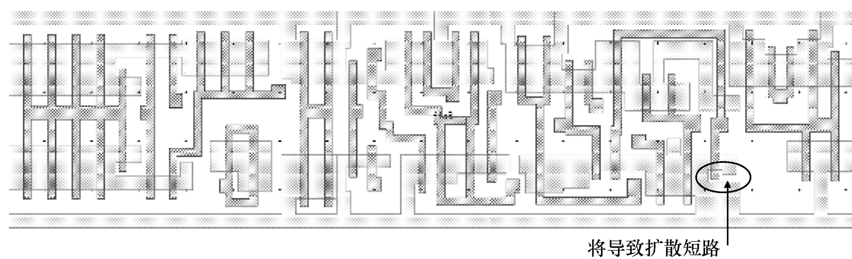


图 10-8　图 10-7 中触发器版图

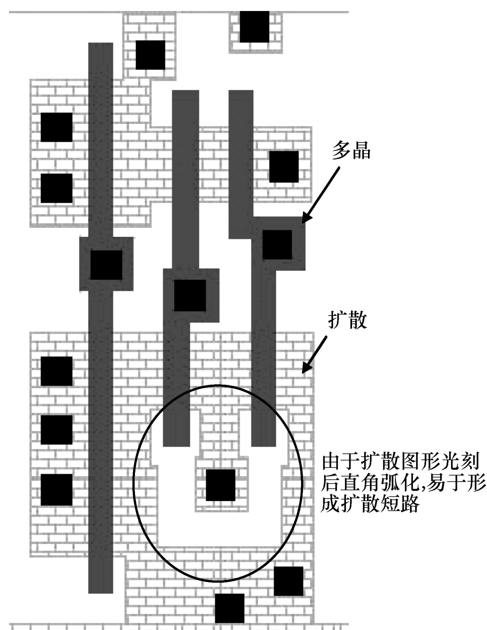


图 10-9　形状较差的扩散槽

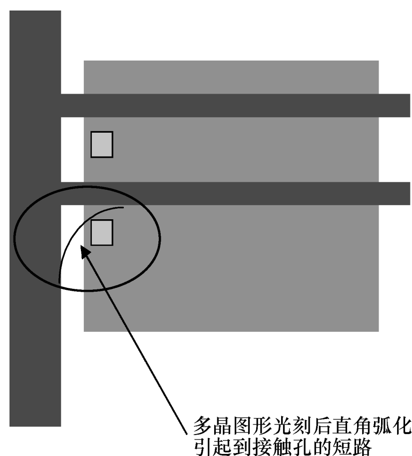


图 10-10　接触太靠近扩散边界的版图

在纳米 CMOS 技术中，即使对版图——光学相互影响方面经验非常丰富的设计者而言，光学仿真也是必需的。这将有助于关键层的版图设计以及接触孔、多晶、扩散和通孔的合适布局。

在一些敏感的情况下，甚至接触孔和通孔的金属化覆盖都会成为问题，因为金属最小宽度和间距变得更小了，因此在 90nm 节点中要求采用 OPC 和相移掩膜版（PSM）技术。基本原则是保持多边图形简单，不要采用复杂的急弯，并且在不增大单元面积的情况下使多晶转弯尽可能远离扩散边界。

10.3　全局布线 DFM

在第 11 章中，我们将研究采用在制造中易于实现的构造时钟布线技术来减少波动以及实施修正。本节我们将介绍能改善其他全局布线特性的技术以及它们对成品率的影响。

即使不使用低 k 介质，铜互连工艺仍然是纳米 CMOS 芯片制造中的主要难点。首要的也是最重要的是芯片上的互连密度。理想的方法是布线器能在整个芯片上实现均匀的互连密度。实际工作中，布线器仍然不具备这种能力。目前已出现了一些工具有助于实现这样的密度要求。虽然尚不完美，至少正在向均匀性目标这一正确方向前进。结合金属填充和开槽等工作，可以实现金属密度

的高度一致。

由最窄互连组成的较长平行走线是导致成品率损失的另一种机制。这种成品率损失的原因是由于漂洗期间光致抗蚀剂侧墙上的毛细现象导致光致抗蚀剂的分解。这种毛细力将光致抗蚀剂的侧墙拉到一起，特别是当一个槽被表面活化物质完全填满而邻近的槽为部分填满时。这种对侧墙作用力的差异引起了光致抗蚀剂侧墙的崩塌。

如果布线器能根据空间情况将互连之间的间距扩展开，将减小引起短路和开路的外部微观材料对成品率的影响，也能减少光致抗蚀剂分解产生的缺陷对成品率的影响。即使是几微米的互连扩展也能对成品率产生重要的影响。扩展互连的另一个优点是能改善性能和信号完整性。图 10-11 是互连扩展布线的一个实例。

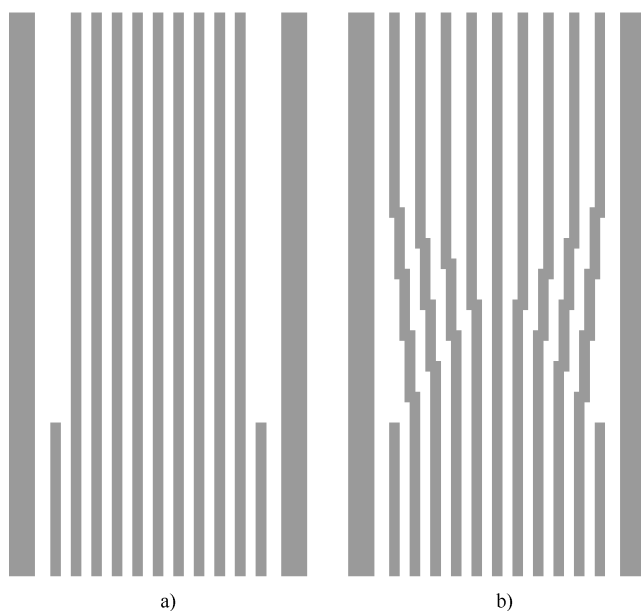


图 10-11　a）传统的布线　b）互连扩展布线

10.4　模拟电路的 DFM

纳米 CMOS 工艺中以数字为中心的优化，给模拟电路设计带来很大的困难。为了确保能成功制造模拟电路，必须考虑相关的关键问题。必须尽早考虑的一个问题是，是否要针对模拟电路开发特殊的设计规则来检查，以保证在存在波动的情况下可实现模拟电路的重复生产。一些应该被考虑的规则实例如图 10-12 所

示。该图总结了数字和模拟器件之间的版图差异。模拟器件很少有过于严格的设计规则，这会使器件芯片面积更大。这些过于严格的规则具有的优点是模拟器件比数字单元对工艺波动的敏感性更小。掩膜对准程度会严重影响模拟单元的性能。使套刻误差对模拟电路的影响最小化的技术将在第 11 章深入讨论。增加间距和覆盖要求可以减小有效栅长的分散性，这是由于多晶图形光刻后直角弧化能减少因为多晶掩膜未对准所产生的变化，如图 10-12 所示。相似地，增加接触孔与栅以及接触孔到扩散区的间距，在接触孔掩膜出现偏移时，也能减小电阻的分散性以及接触孔与栅之间电容的分散性。避免使用最小栅长也将减小模拟器件的分散性。选择栅长时必须多加小心，因为大多数的工艺都是按照数字电路的要求进行调整。为了达到数字电路的高密度，一般基于最小沟道长度确定节距。由此确定的节距，对模拟电路模块来说，很可能波动性较大。如果没有正确地应用 OPC 算法，也可以完成版图设计，但是却会产生成品率问题。强烈建议模拟电路设计人员与集成电路制造方或生产线协商，以确定作为长度函数的波动性，从而能避免使用节距和沟道长度落入“禁区”。为了更好控制 poly-CD，在设计中添加亚分辨率辅助图形（SRAF 或散射条），形成节距“禁区”。纳米 CMOS 技术光学方面的内容和这些效应的详细解释可参见第 3 章。为了更好地实现poly-CD 控制以及使 V_{th}的波动最小，使所有的模拟晶体管同向排列非常重要，详细内容见第 11 章。

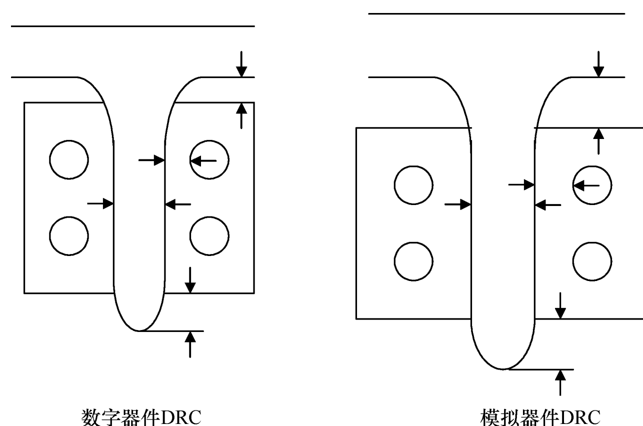


图 10-12　模拟精度规则

有些模拟电路，特别是锁相环（PLL），使用大量的电容作为去耦电容和环形滤波器电容。这些密集的电容阵列增加了扩散区密度，使得浅槽隔离（STI）刻蚀后很难清理扩散区上的氮化层。因此必须将这些阵列分开，以保持扩散区密度低于制造方设定的工艺阈值之下。

10.5 一些基本规则

(1) 只要可能，要避免采用最小间距和最小宽度的互连，以使信号互连的腐蚀变形最小，这些变形会增加电阻并使时序退化，而这是通常工具分析不了的。

(2) 宽的互连要求更宽的间距，因为宽线侧面更容易塌陷，从而引起短路。同时宽线侧墙斜坡也更大，会导致与邻近互连的短路。

(3) 扩散图形光刻后直角弧化会引起宽度较窄的器件的尺寸变化，变化程度与版图设计情况密切相关。如果需要采用较小尺寸晶体管，应该采用不包括“哑铃”形的扩散图形（请参见图 11-17）。多边形形状越简单，工艺处理越容易，并且 OPC 也更简单。

(4) STI 应力将引起迁移率退化，SPICE 仿真中必须包括这一问题。一个更好的解决方法是在设计时尽可能排除应力。第 11 章中将讨论使这些效应远离版图的一些策略。

(5) N_{well}邻近效应会引起 NMOS 发生 50mV 的 V_{th}漂移，对 PMOS 会引起 20mV 的 V_{th}漂移（请参见图 11-23a）。必须注意匹配晶体管的放置，应该使它们与阱的方向及间距一致。

(6) 限制版图设计中的自由度。例如，使所有的晶体管以同样的方式排列可以极大地改善工艺控制和优化。即使逻辑晶体管不能完全采用相同的方向，应该明确，对所有的关键多晶和存储器器件来说，多晶方向对准非常重要。

(7) 设计一致性和使用平铺的器件保证器件的一致性，有助于器件匹配。

(8) 约束多晶节距并且使用“哑元”器件以保证期望的邻近，将使光刻步骤更容易并且可以获得更好的 poly-CD 控制。使用 SRAF 要求多晶节距约束。另一个优点是注入多晶的邻近效应更均匀，这会使波动更小。

(9) 关键版图的对称和精度规则的使用将有助于确保端帽有足够的交叠（详细讨论见第 11 章）。

(10) 多接触孔和通孔的使用对成品率有重大的影响。

(11) 使用更加结构化的设计方法，禁止使用随机性的版图图形。10.2 节已介绍了几个实例，说明随机性版图图形是如何引起严重的成品率问题。

(12) 应该尽可能维持整个芯片上多边形密度的均匀性，需要时可以使用工具提供帮助。需要时填充金属和开槽；应该尽量采用互连扩展这一密度归一化技术。分割电容阵列以减小扩散区密度。

(13) 对模拟单元应使用精度或模拟设计规则。

10.6 总结

为了可制造性和更好的成品率，本章讨论了物理设计方面的因素。第 11 章将讨论电路设计方面的问题，包括一些会加剧工艺波动的物理设计风格及其对 DFM 和电路性能的影响。生产成品率对产品和公司的成功非常重要，而不应该认为只是制造工程师的责任。应用好的设计实践，特别是在纳米 CMOS 技术中，将可以通过减少寄生来改善成品率和芯片性能。正如我们在案例研究中看到的，设计对纳米 CMOS 芯片可制造性和成品率的影响极大[5]。以制造和成品率为对象，针对波动性的鲁棒性电路和物理设计也能间接地导致工艺开销大幅降低并获得更好的性能。

参 考 文 献

[1] Future of semiconductor manufacturing, workshop, *IEEE International Electron Devices Meeting*, 2002.

[2] M. Orshansky, Computer-aided design for manufacturability, University of California–Berkeley, 2002.

[3] B. E. Stine, D. S. Boning, J. E. Chung, D. J. Ciplickas, and J. K. Kibarian, Simulating the impact of pattern-dependent poly-CD variation on circuit performance, *IEEE Trans. Semicond. Manuf.*, Vol. 11, No. 4, Nov. 1998.

[4] F. Schellenberg, Sub-wavelength lithography using OPC, *Semiconductor Fabtech*, 9th ed., MAR 1999.

[5] R. Radojcic, Old rules no longer apply: what’s yield got to do with IC design? *EETiMES*, 2003.

第 11 章　针对波动性的设计

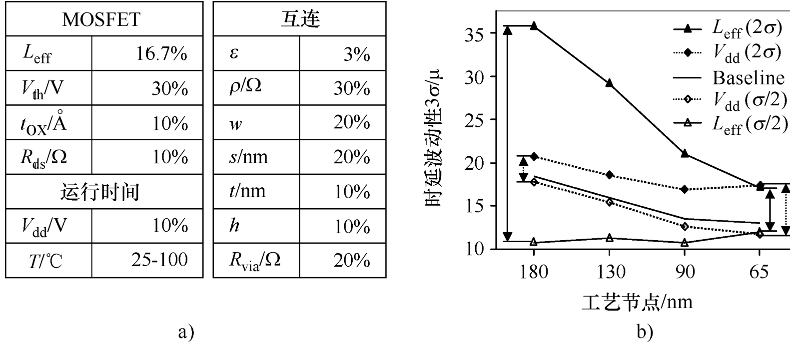
11.1　波动性对未来设计的影响

在过去的几十年中，硅工艺尺寸的快速缩小使集成电路（IC）设计的巨大成功成为可能，并且使得单个芯片中能够集成超过百万个晶体管。然而，当工艺尺寸持续缩小时，芯片制造的精确控制使纳米时代的制造变得日益困难与昂贵。当工艺尺寸继续缩小时，硅工艺，如光刻、氧化、离子注入和化学机械抛光（CMP）等承受着更严酷的波动性影响。另外，工作期间环境的起伏（例如 V_{dd}中的 L（di/dt）噪声和温度变化）的影响也随着芯片工作频率和功耗的迅速增加而更加严重[1-3]，最终使得电路性能出现了更大范围的波动，导致下一代工艺成品率的衰退增加。在先进的 IC 设计中，电路的鲁棒性成为障碍，并且要求工艺工程师和设计工程师努力合作以减轻这种影响。本章将介绍一些用来缓解波动性对设计影响的设计技术。

11.1.1　电路设计中的参数波动

电路参数的波动涉及硅工艺（如有效沟道长度 L_{eff}、阈值电压 V_{th}、金属宽度）或电路工作参数（如信号串扰、电源噪声和温度）与设计标称值的偏离。这些偏离来自芯片制造或运行阶段的电路工作。假定这些偏离服从正态分布，图 11-1a 总结了 130nm 技术节点主要波动源的相对偏差（3σ/均值（μ））。这些数据引自 2002 国际半导体技术路线图（ITRS）[1] 以及一些从科研预测中得到的数据[4-6]。根据 ITRS，在 90nm 或者更小的工艺中会有相似的或者更严重的波动，而工业界预测计的偏差还要大[7]。在这些波动源中，电路时延波动对 L_{eff}、V_{th}、金属尺寸、信号耦合、V_{dd}噪声和温度等最敏感[6]。在当前的工艺中，其他一些参数可能对性能的影响很小（如寄生的源漏电阻 R_{ds}）或它们甚至可以从先进的波动性控制中受益（如介质常数），因此，在变化分析中可以忽略它们的影响。而且，即使是一阶变化，对波动性的影响也会随着工艺的缩小而改变。例如，图 11-1b 所示为 L_{eff}和 V_{dd}控制（即将参数变化减小到 $\sigma/2$ 或使它放宽到 2σ）对规范的关键路径结构性能的影响[6]。关键路径结构是基于 ITRS，并且这些预测是通过使用 BPTM 器件和互连模型经过 SPICE 仿真获得的[4,5]。由于速度饱和，L_{eff}对更短沟道的晶体管的影响很小，而 V_{dd}的重要性随着沟道长度的减小而增加[6]。

注意，在前几代工艺中，电路性能的波动性中，晶体管和栅级的波动占主导地位，但是最近的工艺尺寸缩小已经导致片上互连参数产生了更大的波动，包括互连尺寸、电阻率（ρ）、介电常数（ε）和通孔电阻（R_{via}）[8]。这些互连波动与晶体管的波动无关，并且不同层次互连之间也基本相互无关，导致基于工艺角模型对所有波动性的分析变得极其复杂。



MOSFET		互连	
L_{eff}	16.7%	ε	3%
V_{th}/V	30%	ρ/Ω	30%
t_{OX}/Å	10%	w	20%
R_{ds}/Ω	10%	s/nm	20%
运行时间		t/nm	10%
V_{dd}/V	10%	h	10%
T/℃	25-100	R_{via}/Ω	20%

图 11-1　电路参数的波动情况及其对时延分散性的影响

a）130nm 技术节点主要波动源的 $3\sigma/\mu$ 比值　b）L_{eff}和 V_{dd}对时延的影响

参数波动具有空间分布的特性，因此通常采用芯片内（即 intradie）或芯片间（即 interdie）来表征。芯片间的波动是指在同一个芯片内波动相同，因此对同一个芯片内的每个元件具有同样的影响，而在整个晶圆范围呈现随机分布。这种波动决定了芯片上每个参数的标称值，并且使得这些标称值在晶圆上各芯片之间以及在不同晶圆之间互不相同。目前的工艺中，特征尺寸总的波动中，芯片间的波动大约占特征尺寸总波动的 50%[9]。芯片间的波动基本与设计无关，而与设备的特性、晶圆的放置、工艺温度等因素有关[9]。

芯片内的波动表现在沿着芯片长度方向参数的波动。在以前的工艺中，它的影响可以忽略，但是在纳米时代，它达到与芯片间的波动可比拟的程度，并且在某些情况下甚至远大于芯片间的波动[7]。对关键路径时延的波动性来说，芯片内的波动直接影响着时延均值，而芯片间的波动决定了时延分散性的方差[9]。芯片内的波动可以进一步分为两部分：系统的和随机的。系统波动可以在制造前进行预测，例如版图对沟道长度分散性的影响。成功的工艺缩小依赖于工艺和设计阶段对系统波动因素的有效补偿。相反，随机波动是由于半导体工艺本身的不可预测性引起的。随机波动的例子包括沟道掺杂、栅氧厚度和介电常数等的波动。运行阶段的一些参数波动，如 V_{dd}噪声，也被认为是随机波动，由于精确预测它们的影响极端困难，因为我们不能对随机现象进行补偿，这种波动可能最终对具有满意成品率的纳米 CMOS 电路设计提出巨

大的挑战。

对给定的工作温度，L_{eff}、V_{th}和 V_{dd}的随机波动是逻辑门中参数分散性的最主要来源。L_{eff}和 V_{dd}的波动彼此相对独立，而 V_{th}值同时与 L_{eff}、V_{dd}值以及晶体管尺寸有很强的关系。这是因为短沟 MOSFET 的 V_{th}标称值直接受 DIBL 效应的影响，而 DIBL 效应是 L_{eff}和 V_{dd}的函数，描述 V_{th}分散性程度的参数 $\sigma_{V_{th}}$则依赖于晶体管尺寸，并且主要受到沟道掺杂波动的影响。下式表示的 V_{th}分散性和晶体管尺寸之间的关系适用于纳米时代[10,11]（同时请参考 11.3 节）：

$$\sigma_{V_{th}} \propto (W_{eff}L_{eff})^{-1/2} \tag{11.1}$$

为了针对波动性进行正确的设计和优化，必须考虑这些关系；在电路级，可以利用它们之间的关系，通过调整 V_{dd}、V_{th}和晶体管尺寸，获得性能、功耗和分散性之间的折衷[11]。

11.1.2　对电路性能的影响

研究发现，电路参数波动性的增加引起电路性能参数呈现更大的分散性[6]，进而导致芯片的成品率退化。成品率是指所有电路中传输时延满足关键时延要求的那一部分所占的比例。图 11-2a 所示为 130nm 工艺的 4 位加法器采用 Monte Carlo 仿真得到的时延特性参数数据分布直方图。使用图 11-1a 中给出的分散性数据，在标称偏置条件下（$V_{dd}=1.2\text{V}$），描述性能分散性的 $3\sigma/\mu$ 高达 15%。另外，如图所示，为了减小功耗而使 V_{dd}减小到 0.5V 时，分散性变差，$3\sigma/\mu$ 达到 45%（见图 11-2a）。注意，在低 V_{dd}时，性能分布变得不对称了，这是由于 CMOS 电路对偏置条件的非线性响应[10]。这种情况下，采用对数分布模型来描述其统计特性，因为它可以获得比普遍使用的正态分布模型有更好数据拟合效果，特别是平均值的提取[12]。除了对分散性产生不好的负面影响，参数波动也使功耗问题更加突出，特别是泄漏功耗方面。图 11-2b 汇总了 4KB SRAM 芯片的实验结果。与没有考虑波动时仿真得到的泄漏值相比，测量得到的泄漏电流（I_{leak}）在较大的 V_{dd}范围内成指数上升，进一步威胁着 SRAM 的正常功能，并且增加了图形对波动的敏感性所引起的失效。泄漏的剧烈增加是由于晶体管 L_{eff}更短和 V_{th}更低引起的：由于漏感应势垒降低（DIBL）效应，使得那些 L_{eff}值更小的晶体管的 V_{th}退化更加严重，导致 I_{leak}与 V_{th}之间成指数关系。因此，它们对波动非常敏感[13,14]。不幸的是，功耗已经成为当前高性能设计的主要障碍之一；波动的增加进一步引起了功耗问题，并且因此增强了这种障碍。因此实现鲁棒性设计的技术就成为未来 IC 成功的关键要求。

如图 11-2a 所示，分散性随着 V_{dd}的降低而进一步严重，这意味着电路成品率随着功耗的减小而衰退。然而，这种现象并不是只在调整 V_{dd}时发生；当对 V_{dd}

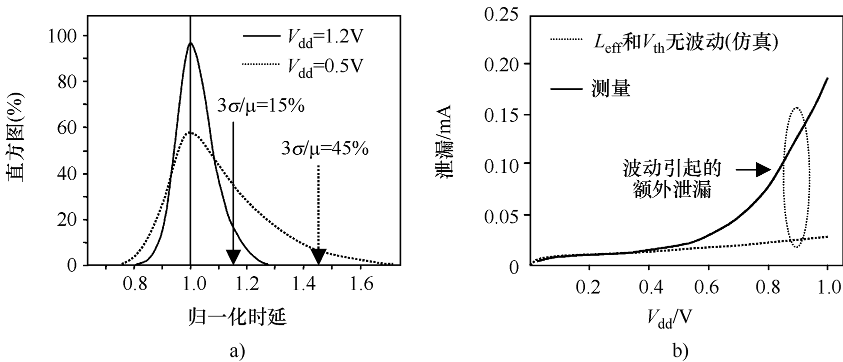


图 11-2　130nm 技术中波动对时延参数分散性和泄漏功耗的影响

a）4 位加法器的 Monte Carlo 仿真结果　b）测量得到的 4KB SRAM 泄漏

和晶体管尺寸进行调整时，它也会发生。业已表明，为了平衡功耗减小和性能退化之间关系，最有效的技术之一是实现 V_{dd}、V_{th}和晶体管尺寸之间的协调，并且在这三个参数之间进行折衷。然而，在节省功耗的优化期间，时延分散性的增加速度与标称时延相似，因此优化时导致成品率会减小[10,12,15]。对通过时延优化而确定的具有一定大小的 130nm 工艺反相器链，图 11-3 所示的转换能量和成品率与 V_{dd}和 V_{th}之间的函数关系证明了这一结论[12]。如实验和仿真所示，只要能量和时延不超过它们各自的约束，为了改善成品率，希望使用更高的 V_{dd}和更低的 V_{th}[10,12]。而且，图 11-3 还表明，随着 V_{dd}的减小，转换能量急剧减小，而成品率却以相对较低的速率衰退。这种关系意味着电路级的能量——成品率折衷对低功耗设计来说作用是明显的：在可接受范围内的成品率牺牲可以实现明显的能耗减小。

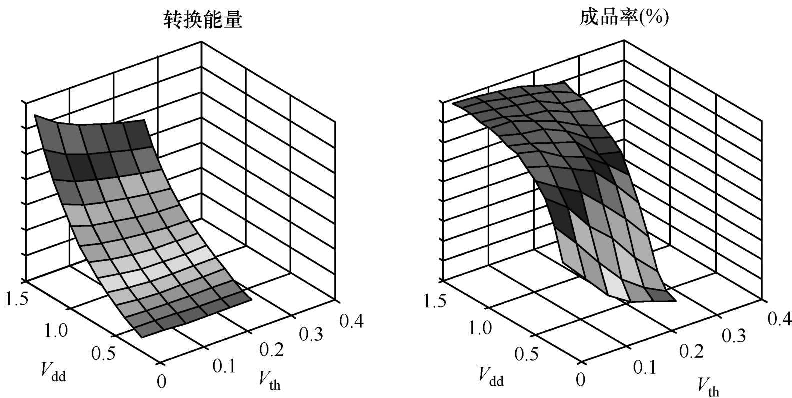


图 11-3　功耗减小引起的成品率衰退

11.2　减轻波动影响的策略

11.2.1　使偏斜最小化的时钟分布策略

当微处理器时钟频率增加到 3GHz 以上时，时钟偏斜逐渐成为决定时钟周期的重要因素，因此需要使时钟偏斜最小化，才能进一步提高工作频率。而工艺波动已成为影响时钟偏斜的重要因素[16]。这种影响不仅体现在锁相环（PLL）设计中，而且体现在时钟网络设计中。在时钟分布网络设计中，一种减小偏斜的非常有效的方法是使用时钟网格，如 AlphaEV6（21264）微处理器中就采用这一技术[17]。不过，当芯片规模太大时，这种方法就失效了。当缩短的网格的 *RC* 时延等于或大于期望的时钟偏斜时，采用网格将不能改善偏斜；而其寄生电容将增加时钟网络的动态功耗。

随着工艺尺寸的缩小，每个芯片具有的功能也在不断增加，因此性能最高的处理器芯片的尺寸并没有随着工艺尺寸的缩小而减小；在某些情况下芯片尺寸反而增大了。虽然随着工艺尺寸的缩小，门时延在不断减小（见第 1 章），而互连时延却仍在增加。如果互连尺寸不随之缩小，则时延更长（见第 1 章，图 1-7 和图 1-8[19]）。因此，为了使 *RC* 时延与门时延相当，互连的长度必须越来越短（参见图 1-8c）。这迫使设计者使用更密的网格以改善偏斜，获得比非网格化的时钟分布网络更好的偏斜。这最终会导致更高的时钟动态功耗。EV6 处理器上的网格占用了相当多的功耗，其中时钟分布网络的总功耗约占芯片总功耗的 40%。全局网格电容是 2.5nF，而在包括锁存器的局部分布中，主要的网格电容是 3nF 和 6nF[22]。而网格本身消耗了大约芯片总功耗的 19%[17]。网格越密，功耗越高。在低功耗设计中，更倾向于采用无网格的时钟树，而不是网格化的时钟分布网络。这是因为网格化的时钟分布系统需要相对大的功耗实现较小的偏斜[18]。

平衡 H 型树分布的使用正变得更加普遍，它不仅可以获得更好的功耗性能，而且可以减小网格化时钟分布的增益。正如我们看到的，在纳米 CMOS 时代，为了使 *RC* 时延等于门时延，要求互连的长度不断缩小。然而，H 型树分布系统也受工艺波动偏斜的影响，这就要求实现负载平衡以获得小偏斜。这是因为 H 型树的负载电容与互连电容近乎相等。因此，H 型树情况下的负载电容比时钟总电容更大。网格时钟系统中，互连电容占主导地位，因此对负载的不平衡很宽容，不要求负载平衡来保证低偏斜。

为了使工艺波动引起的偏斜最小，需要使用非最小沟长器件作为时钟驱动。这必须在面积、功耗和偏斜之间进行折衷。为此，需要理解增加 halo 注入晶体管沟道长度的影响，以及确定按照什么沟道长度来设置工艺才能使多晶层的特征

尺寸（CD）控制最佳。所有的制造商都愿意优化最小的多晶尺寸来实现最好的CD控制。他们可能不总是成功；因此我们需要获得那些来自制造商的数据，在最低的CD变化点设定沟道长度或能提供最低CD变化的最小沟道长度。如前面描述的，尽管沟道较长的器件中绝对变化可能更大，但是以百分比表示的CD相对变化甚至比最好点更低。使用较长沟道长度的问题是增加了面积、功耗以及锁相环（PLL）所需的缓冲级数。由于电源电压和器件的变化，PLL要求的缓冲级数越多，引入的偏斜越大。

在时钟缓冲器中使用去耦电容，可以很好地抑制由翻转活动引起的电源电压降低。另外，当时钟缓冲器使用集成去耦电容时，通常围绕在缓冲器周围的，是一个由非翻转器件形成的“环形低功率区”，因此减小了时钟缓冲器周围的功率密度。这就减小了对功率分布的需求，并且当时钟缓冲器位于高功耗电路（例如执行单元）附近时，可以提供与热源之间的隔离。由于电源电压下降的幅度减小，位于芯片上“较冷”与“较热”区域的时钟缓冲器之间的温度差降低，最终的时钟设计将具有更低的偏斜。为了改善性能而将芯片上温度高的区域最小化的做法也是很好的设计方式。高功耗模块需要通过电源供电来维持边缘变化率；否则将会限制模块本身和整个芯片的性能。因此，插入去耦电容将高功耗模块分割开，不仅可以保持电源电压的完整度，而且可以减小此模块所在区域的功率密度。由于温度较高区域的缓冲器与温度较低区域的缓冲器之间的温差，当时钟缓冲器被放在温度高的区域时，会增加时钟偏斜，因此插入去耦电容将高功耗模块分割开就更加重要。

时钟缓冲器的版图需要像模拟版图一样认真对待。由于水平和垂直CD的差异，必须在芯片上以同一方向放置时钟缓冲器的版图。130nm节点中，除了掩膜本身的误差，水平和垂直方向多边形之间会有2nm的变化，并且如果有的话，对90nm节点这个值不会有明显减少。水平和垂直方向的变化是由于在水平方向边缘打开和关闭写入电子束与垂直方向存在差异。刻蚀后最终像的CD变化将进一步变差约二到三倍。不同时间在器件中进行halo注入也在水平和垂直方向引入了V_{th}变化。

时钟缓冲器一般非常巨大，因此将它们划分成较小的晶体管是很重要的，这样就可以减小长而窄的多晶互连电阻的波动。器件应该总是被分割为偶数个数。当晶体管被拆分时，对未对准情况就不很敏感了（见图11-4）。其他的未对准效应将在本章后面部分进行讨论。

对所有的时钟缓冲器，应该仅使用一种版图，并且应该尽可能遵守模拟版图规则（见11.2.3节中针对模拟电路波动性的策略）。可以通过使NMOS浅槽隔离（STI）应力下的迁移率衰退最小来使得器件的驱动强度最大化（见图4-4）。采用哑元晶体管可以实现这一目标，并且可以改善由多晶刻蚀器件的微负载效应、光刻效应以及多晶栅侧墙（见图11-23b）上的注入散射引起的邻近效应等带来的CD波动。

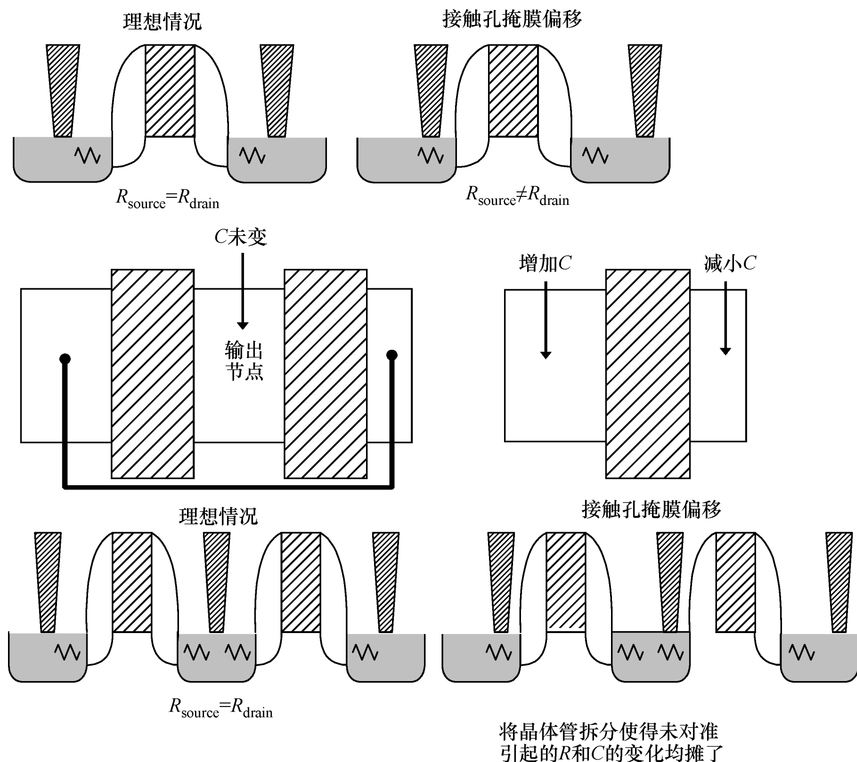


图 11-4　未对准对寄生 R 和 C 的影响（解决方法是拆分晶体管）

GHz 级的芯片中，对时钟互连的屏蔽是必须的。除了提供容性屏蔽，它们也通过对时钟和其他攻击信号提供信号回路而起着感性屏蔽的作用。因为屏蔽是采用手动方式放置的，可以在掩膜数据预准备阶段采取更准确的提取，所以可以在版图数据被送到掩膜写入之前，保证结构设计的正确性。在掩膜数据预准备阶段，在金属密度低于约 20% 的区域进行金属填充。在增加金属填充以使时钟和屏蔽周围以及时钟布线层之上或之下的金属密度正常化后，屏蔽就将边缘电容波动的影响最小化了。遵守宽互连的间距和宽度规则，以使成品率损失以及由化学机械抛光（CMP）效应如腐蚀和凹陷（参见第 2 章，图 2-26）等引起的电阻波动最小。此外，放置屏蔽有助于减轻由 CMP 效应带来的电阻波动。确保时钟线之上和之下的互连密度均匀非常重要[21]。非均匀互连密度会导致层间介质（ILD）厚度变化，从而引起线时延变化和时钟脉冲偏斜（见图 11-15）[20]。互连密集区域的 ILD（t_1）较厚而互连稀疏区域的 ILD（t_2）较薄。CMP 将会消除一部分但不能完全消除严重的 ILD 厚度变化。一些工具可以通过填充和开槽来帮助实现图形密度的规范化。还有一些工具是与布线器一起联合工作，通过改变互连的疏密来改善图形密度，但是它们不能完全解决密度问题。它们需要与金属填充

一起使用来使密度规范化[21]。

在时钟树中，铜互连会受制于由亚波长光刻、腐蚀和凹陷引起的变形，而这些改变了互连的宽度和厚度，因此改变了 *RC* 时延，引起较高的偏斜[25]。对电路设计者而言，理解其版图设计对这些效应的影响并且与工艺工程师一起工作来确保这些新的恶化的物理效应不会严重影响各种时钟树分支的时钟分布网络互连时延，是非常重要的。对腐蚀和凹陷，使用一些较窄的互连会更有效，其效果与制造密切相关。对于其他效应，在较宽的时钟分布互连上开槽会更好。使用几条较窄互连的优点是具有更大的表面积，有利于获得更好的高频电阻率。改善程度的高低取决于频率和趋肤深度。对大多数生产线，使用几条互连形成较宽的互连还可以更好地解决腐蚀和凹陷的问题，因为互连之间介质的宽度大于槽的宽度，因此可以更好地抗腐蚀。这是结构校正方法的一部分，并且将是未来纳米 CMOS 时代设计的规范。

如果没有考虑由扩散和多晶图形的直角外扩效应引起的尺寸变化，就会影响时钟偏斜[23]。如果时钟互连必须驱动更大数量的这种晶体管，特别在阵列设计中，问题将更加严重。在采用时钟控制的灵敏放大器中也会存在这一问题，其中同一条时钟线被连接到大量的相似器件上，而这可能是 128 个甚至是 256 个例化元件。画出的器件版图可能很小，但是由于扩散的图形的直角外扩效应，加工出的器件宽度会增加多达 25%（增加程度与版图有关），导致从驱动器看进去的负载就随之增加（见图 11-18）。为了使这个效应最小，在器件中必须尽可能不采用“哑铃”形状的扩散。较好的设计准则是器件使用最小宽度而不采用“哑铃”形状的扩散，并且对负载进行精心设计以避免工艺加工后因负载增大带来的“意外”。当我们深入到亚波长光刻时，这个问题会变得更加严重。如果使用允许的最小尺寸的晶体管的目的是缩小版图面积，无“哑铃”形状扩散的晶体管可能会占用相同或更小的版图面积，并且在波动性和驱动方面性能更好。

波动的另一个来源是晶体管被连接到时钟的方式。由于晶体管连接到时钟的方式不同，对时钟呈现的电容也就不同[22]。通过传输管来馈送时钟（见图11-5）是一种较差的结构。而这种结构是一些高速缓冲存储器设计中普遍使用的，其目的是减小解码器 NMOS 堆栈的堆栈高度从而改善速度。然而，它会给时钟引入一个依赖于寻址模式的负载。如果寻址模式打开传输管 M1 到 M*n* 中的任何一个，时钟将会看到一个比传输门关闭时更高的负载。时钟负载还进一步依赖于解码器 N 型树上的数据模式。这就产生了一个依赖于寻址模式的时钟脉冲偏斜，因此应该避免这样的设计。

11.2.2 针对波动性的 SRAM 技术

位单元和读出放大器是 SRAM 中最重要的两个部分，也是对工艺波动最敏

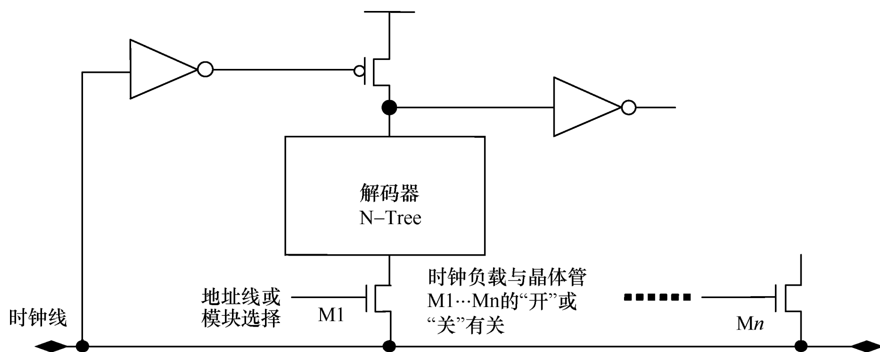


图 11-5　时钟作为解码器的逻辑输入

感的部分。对位单元，其尺寸较小是它易受工艺波动、特别是 V_{th}波动影响的主要原因。离子注入的统计波动、穿过栅进入沟道的掺杂通道（见图 11-6）、多晶和扩散的 CD 波动以及通过隔离氧化层的掺杂损失是 V_{th}波动的主要诱因[25,26]。由于影响 V_{th}的尺寸波动，以及阈值电压的滚降特性和反向短沟（Reverse Short-Channel，RSC）效应，包括窄沟效应，使得面积小于 $3\mu m^2$的位单元对 V_{th}的波动特别敏感。尺寸波动是工艺波动以及工艺与版图相互作用的结果。位单元版图对这些效应有极大的影响，将在后面分析说明。随着技术的进一步发展，特征尺寸缩小到 100nm 以下时，这些效应将进一步加剧。同时，为了达到高速度及高密度的目的，大多数存储器都采用人工布线，因此了解在设计中如何将这些效应的影响减小到最小至关重要。另一方面，由于依靠人工设计，一个很重要的问题是确保这样的设计也能适合于未来技术节点中的尺寸缩小，从而使修改费用降到最低。

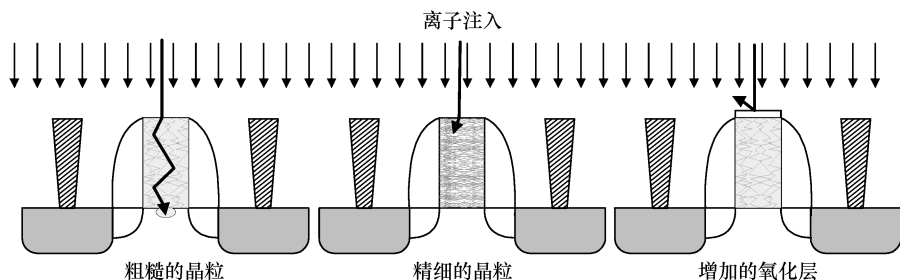


图 11-6　掺杂沟道

位单元设计中的设计问题　多晶与扩散 CD 以及注入波动是导致两个交叉耦合驱动器之间驱动失配的主要原因。图 11-7 中，M1 和 M3 构成了其中一个驱动器，而 M4 和 M6 组成了另一个驱动器。六管位单元中的匹配晶体管对是 M1/M4，M3/M6 和 M2/M5。为确保单端口位单元出现侵入式读取（读取/干扰）时单元的稳定性，传输（M2 或 M5）与下拉（M1 或 M4）NMOS 之比必须满足一定的值

(1.8~2.2)。增大比率会带来更大的代价——单元尺寸的增大。位单元设计要折衷考虑多方面因素，必须在最高成品率、最小单元尺寸之间实现最佳性能平衡。工艺波动以及版图与工艺之间的相互作用会影响器件的匹配以及关键的比率关系，如上所述，会进而导致工艺偏移窗口的裕量和容限的减小。如果在设计和版图阶段没有考虑本节描述的问题，那么设计将会存在成品率降低和性能变差的风险。

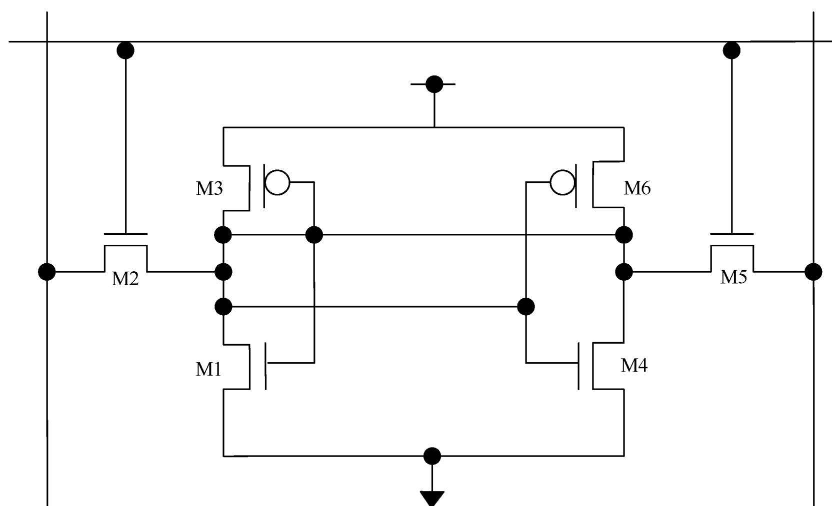


图 11-7　位单元电路图

当位单元的节点电容随着工艺缩小而减小时，在存取期间激活的字线（WL）会导致位单元中由 M1 和 M3 的漏或 M4 与 M6 的漏形成的存储节点上严重的差分噪声耦合。这个噪声的起因是由于存取期间 WL 冲向高电平。由于位单元中的一个节点是高电平而另一个节点是低电平，并且位线（BL）被预充电到高电平，因此其中一个传输晶体管处于饱和模式而另一个处于关态。处于饱和的晶体管会将其栅电容的 2/3 耦合到源节点上，在这种情况下源节点是位单元的“低”存储节点。与位单元高节点连接的传输晶体管源和漏处于 V_{dd}电平。由于 WL 在存取期间从低变到高，因此，对于连接到单元中高节点的传输门，栅源电压（V_{gs}）的初始值为负值。当 WL 为高电平时，这个传输晶体管的 V_{gs}为零，这意味着晶体管仍然处于截止态。当 WL 由低变到高时，对截止的传输晶体管，仅有交叠电容（C_{gd}）耦合到其高电平节点上。因为 C_{gd}约是 2/3 的栅电容的 1/5，因此，低存储节点接收的耦合脉冲将比高存储节点的强。这样，由于 WL 耦合到低存储节点和高存储节点的耦合电容的失配，在存储节点产生了耦合差分噪声。除此之外，单元电流流入单元的低节点，也会引起低节点电位的升高。因此，只

考虑静态噪声容限无法保证设计出性能稳定的单元，除非设计者可以考虑所有的动态条件，上面已经讨论了其中的一个。

与硅局部氧化隔离（LOCOS）情况相比，采用 STI 的晶体管，ΔW 值得到了极大的改善。然而，由于位单元中使用的晶体管的宽度非常小（0.1～0.25μm），因此，即使相对较小的 ΔW 值在一些位单元中仍然显得很重要（10%～20%），需要予以重视，并在对位单元晶体管进行建模时需要考虑 ΔW。

图 11-8a 所示为一个采用了基于模型的光学邻近效应修正（MOPC）后的位单元版图，而图 11-8b 是加工后晶圆上的位单元实际图形。如图 11-8a 所示，绘制的位单元版图的多边形拐角是 90°的直角。然而，经过硅工艺处理后，得到的图形与版图图形不太相同。由于亚波长光刻和等离子刻蚀（RIE）造成的失真，拐角变成了圆形，如图 11-8b 所示。工艺处理后的失真可以通过采用适当的 MOPC 技术降低到最小。然而，单采用 OPC 并不能补偿所有的失真，特别是当光刻的发展趋势显示亚波长光学光刻误差随着工艺向下一个节点的发展而逐渐变大（见图 1-5）。我们将针对不同的位单元版图，分析版图对失真水平的影响以及对设计裕量和性能的影响。通过分析可以清楚地看到，设计者们必须越来越多地通过改进版图设计将光刻和刻蚀造成失真的影响降低到最小。

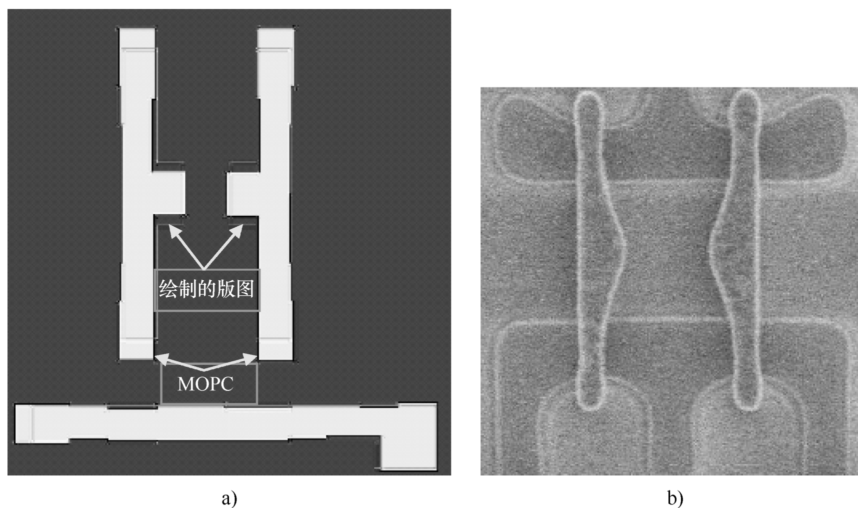


图 11-8　a）采用 MOPC 交叠绘制的位单元版图　b）刻蚀后的位单元图形

由于多晶与扩散未对准，图形失真是导致位单元波动性的另一个原因。一些位单元设计中多晶与扩散区的交叠情况如图 11-9a 所示的那样。这种位单元设计中，考虑到金属 1 形成交叉耦合互连以及多晶栅上接触点的位置，设计中采用了多晶与扩散区不对称的放置方式。由于多晶没有位于扩散区曲线的中间部位，因此对光刻中的套刻未对准非常敏感。接触孔采用如图 11-9c 所示的位置放置是为

了使单元宽度最小。这个设计中的接触点与图 11-8 相比进行了翻转，在图 11-8 中多晶位于扩散曲线的中间部位。实际上，图 11-9c 中为了节省少量的面积而导致扩散曲线上方多晶栅的不对称放置是得不偿失的，因为这使得最终的设计对多晶与扩散之间的套刻未对准非常敏感。如果多晶上的接触点位置与单元中心相距较远，多晶接触将阻碍金属 1 交叉耦合互连，因此必须向位单元的中心移动，以避免短接到交叉耦合金属 1。为此要求将多晶栅移向单元的中部，从而导致扩散区上的多晶位置不对称，如图 11-9a 所示。

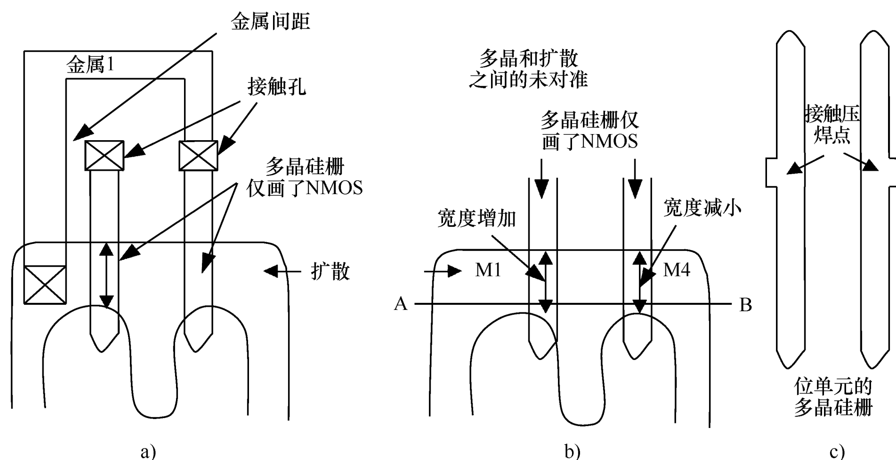


图 11-9　位单元的未对准问题

当存在套刻未对准时，例如在本例中，多晶位置相对扩散区发生右移，导致下拉 NMOS 管 M1 沟道宽度增加而晶体管 M4 沟道宽度则减小。对这种位单元，较好的版图设计如图 11-8 所示，其中多晶位于扩散区曲线的中部，可以使水平方向未对准引起的器件尺寸变化最小甚至消除。唯一的差别是多晶栅上的接触点方位被翻转了，目的是为了增大与交叉耦合 M1 之间的距离，这样就可以使多晶位于扩散曲线上方的对称位置。对这种设计，水平方向的未对准不会引起由扩散上方多晶不对称放置导致的严重失配问题。采用这样的设计，在完全对准的情况下，下拉晶体管可以做到沟道宽度最窄。在出现水平方向套刻未对准时，只是增加了下拉晶体管的沟道宽度从而导致单元比的增加。这也是将多晶与扩散曲线相对对称放置带来的又一个优点。与图 11-9 所示的扩散曲线上方多晶放置不对称的情况不同，在水平方向出现套刻未对准的情况下，这种设计实际上能改善单元的稳定性，而对于多晶与扩散曲线不对称放置的情况，单元稳定性则变差。

另一个位单元版图设计如图 11-10 所示，多晶栅上的接触点在垂直方向上不对称。经过工艺流程后硅片上的位单元图形如图 11-10b 所示，在多晶栅接触点处形成一个喇叭形状区域。由于它们与扩散边界之间的间距不对称，多晶和扩散之间

垂直方向的未对准将导致器件尺寸的变化。如果多晶未对准方向是向下的，由于受多晶上喇叭形状的影响，M1 管的有效沟道长度会增加，与 M4 相比，其驱动能力变差。与此同时，由于受接触孔四周喇叭形扩散区的影响，M2 沟道宽度增加。因为 M2 和 M4 的扩散接触点不同，在发生多晶向下方未对准时，M2 管扩散图形光刻后直角弧化区域更靠近多晶边界，因此导致 M2 和 M4 之间尺寸变化的不对称。这导致了双重的不良后果，即 M2 驱动能力增大而 M1 驱动能力减小。这进一步有效地减小了单元比并且引起了前面描述的匹配晶体管之间驱动能力的失配。

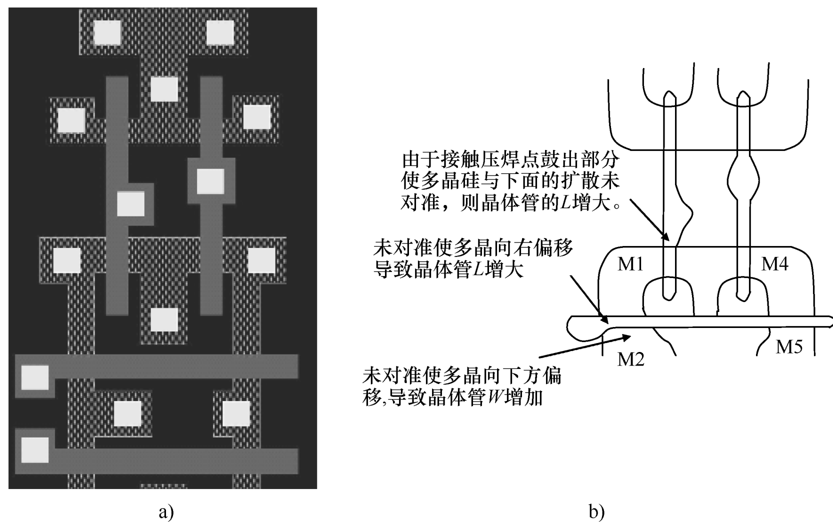


图 11-10　不对称导致的工艺敏感性

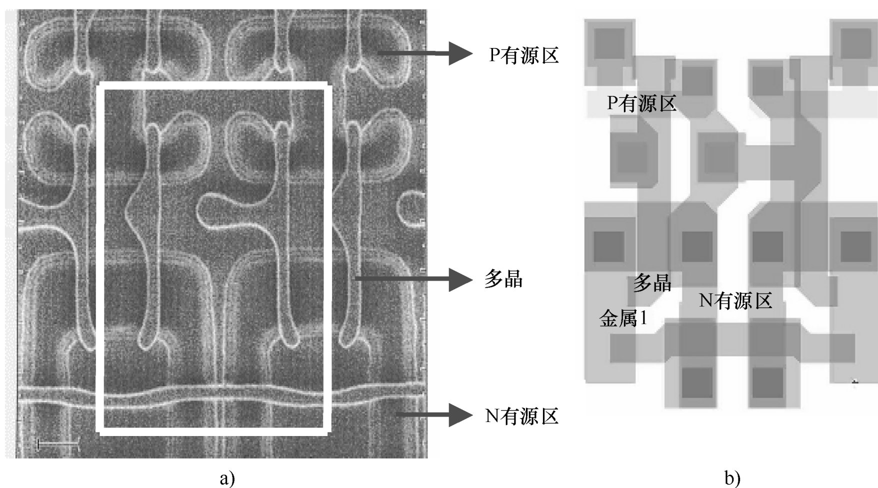


图 11-11　a）光刻和刻蚀后形成的位单元中多晶和扩散图形　b）位单元版图

如果水平方向未对准的情况是相对扩散图形出现多晶向右偏移，由于多晶上的喇叭形状接触点进入到 M2 管的扩散图形，导致 M2 管有效沟道长度增加，使 M2 驱动能力变得比 M5 弱。因此这个设计中的晶体管的匹配对套刻未对准也是敏感的。采用相同的分析方法，可以推断，图 11-11 所示的设计对套刻未对准具有相似的敏感性，这是由于两条多晶栅图形光刻后直角弧化区域以及扩散图形光刻后直角弧化之间存在位置的差异。因此，这个设计对水平和垂直方向的套刻未对准都很敏感。

从工艺加工之前的版图上看不出上面分析的这些影响。但是对于位单元设计者来说，一个非常重要的问题是应该明确，在制造过程中设计将经历哪些类型的失真。只有与制造和工艺工程师的合作，物理设计者才能认识这些效应，从而对版图进行校正以避免这些效应带来的问题。我们将介绍一个位单元设计的例子，说明如何采用相关技术消除前面描述的由工艺失真引起的大部分问题。

纳米 CMOS 领域中将会得到更广泛发展应用的位单元设计如图 11-12a 所示。显影和刻蚀后硅片上的单元如图 11-12c 所示。在该设计中，所有的多晶均同向放置，这样更便于实现多晶 CD 控制、更易于进行光刻及相移掩膜（Phase-Shift Masking，PSM）、以及更好地实现工艺控制[24]。当排列好单元后，所有的晶体管都具有相同的多晶图形，因此多晶邻近问题将被最小化。多晶邻近效应是一种新提出的由于多晶邻近的差异导致注入发生波动而引起性能退化的效应，（见图 11-23b）。这个效应主要是由于邻近区域的其他晶体管的多晶栅对注入杂质的散射引起的。当晶体管的邻近情况变化时，这个效应也会变化。如图11-12a中版图所示，这个设计保证了存储器阵列中的邻近是一致的。

由于该单元不需弯曲形状的扩散图形以及由于接触点和版图对称引起的多晶图形光刻后直角弧化现象最少，因此与图 11-9、图 11-10 和图 11-11 所示的设计相比，对套刻未对准将更不敏感。图 11-12d 显示的是一个相似的单元设计，但是其中扩散区图形还存在少量弯曲。因此，这个单元对套刻未对准引起的尺寸变化就有一些敏感性。改进的单元如图 11-12a 和图 11-12c 所示，其中所有的扩散边界都是直线。只要将传输 NMOS 的长度调整到一个合适的尺寸，就可以实现这一设计。实际上，为了控制纳米 CMOS 电路中单元的泄漏电流，也必须增加传输晶体管的长度。为了同时满足单元驱动能力的需求，可以通过增加晶体管宽度的方式来补偿长度的增加。通过适当调整传输晶体管的宽度，可以使传输晶体管与下拉晶体管 NMOS 宽度比是 1∶1，而两个晶体管的有效 β 比是 1.8∶1 到 2.2∶1 之间的一个合适值，以保证单接口侵入式读出单元的稳定性。这个设计中的扩散区图形不存在弯曲的边缘，因此它是一个对光刻要求不高的设计。另外，由于扩

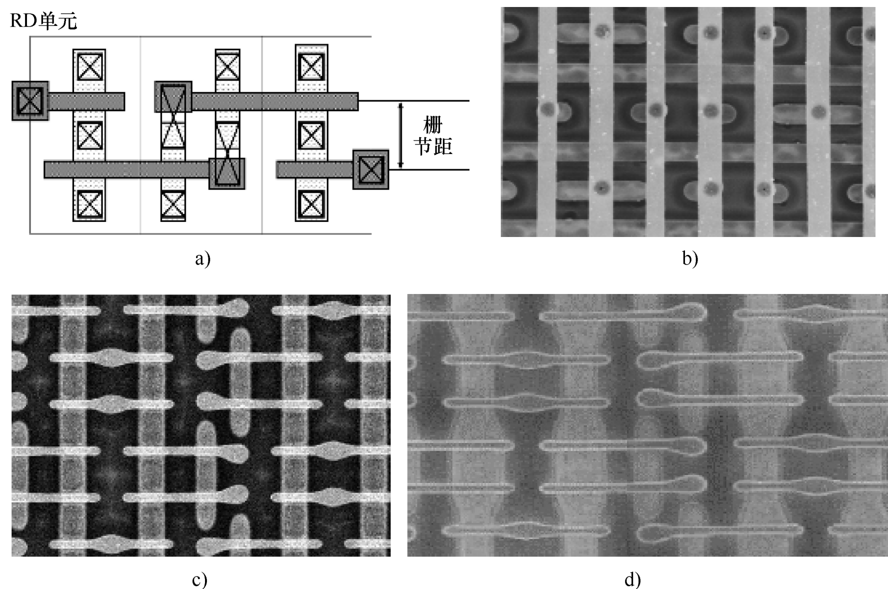


图 11-12　对工艺偏移有最大容限的结构化设计
a）版图　b）单元金属化　c）多晶和扩散　d）多晶和扩散的早期版本
（SEM 和版图由 Treceti/Hitachi 提供）

散区的端头在阵列的末端，就不存在 STI 应力效应。然而，一个重要的问题是要认真处理阵列的末端以减轻阵列末端单元与包括哑元晶体管在内的阵列中心单元之间的差异。

这种位单元设计中多晶的放置方式很好地遵循了纳米 CMOS 设计方法中关键多晶的摆放原则，使得片间多晶 CD 变化最小。注意，无论是多晶、扩散或金属层，其多边形特性的边缘应该尽可能的画为直线。每个层次采用规则的多边形图形都使其更容易进行光刻和工艺控制。保持结构内的图形密度均匀，将使晶体管的邻近效应也是均匀的。其他的工艺控制，如 CMP，也会从均匀的图形密度中受益。未来肯定可以看到更多这样的设计。

这种单元设计还有许多其他的优点，但是它们超出了本书的范围。关于这种单元设计其他优点的详细讨论可参看参考文献［28］和［29］。

设计中光学邻近修正的表征　大多数位单元版图设计在交付掩膜制造之前，都要进行人工光学邻近修正（OPC）。这也是一次修正光学失真的机会。必须合理设置扫描器和光源波长，使修正达到最优，否则将因为过调或者调整不够造成失真。例如，对于图 11-13 所示的实例，多晶条端头的“榔头形”图形太大，使光刻后硅片上多晶互连呈现喇叭状。再加上接触点的喇叭状，使得多晶栅看起来像一个可乐瓶。最终使设计的单元又会对另一个方向（对于本例，是垂直方

向）的光刻未对准很敏感。M1 和 M4 的有效沟道长度与多晶到扩散的光刻对准情况密切相关。如果多晶相对于扩散上移，扩散边界进入到多晶的喇叭状端头，就会导致晶体管有效沟道长度的增加。这就减小了下拉晶体管 M1 和 M4 的驱动能力，因而减小了单元比及其静态噪声容限，导致单元的稳定性变差。因此必须给位单元进行适当的 OPC，以避免对光学邻近效应修正过度或者修正不足。这可能会要求几次重复操作以达到最优补偿，做到不会因为修正过度导致端头呈现喇叭形，也不会因为修正不足导致多晶条端头过分回收。修正过度或者修正不足还会带来其他影响。对 OPC 的详细讨论见第 3 章。如果多晶互连端头回收比较厉害，由于端头覆盖不足，会引起晶体管泄漏。这种欠补偿同时引起单元的稳定性问题以及较高的待机功耗。因为 OPC 对位单元的性能和成品率有重要影响，通常采用人工方式并通过投片进行反复多次的精确调整。

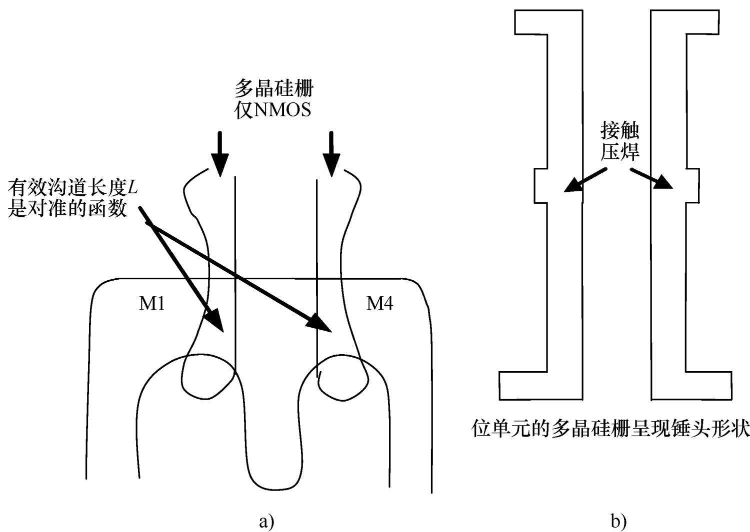


图 11-13　OPC 的过度修正问题

图 11-14b 是一个使用了 OPC 的 90nm 位单元。在 90nm，不能再简单地仅采用榔头形状进行 OPC。应该关注那些复杂的可以用来适当修正光刻和刻蚀失真的 OPC 图形。为了使光刻和刻蚀后的晶体管有适当的沟道长度，可以适当的调整多晶线宽。为了达到适当的校准，需要反复进行并且要同时保证单元的电学参数。关键是对光刻和刻蚀进行正确的修正。可以看到沿着单元的两侧，在多晶层中有亚分辨率辅助图形（SRAF）。因为它们是亚分辨率，所以在光刻中它们不会在硅片上生成图形，但是，当多晶图形非均匀时，特别是在阵列中衬底和阱的接触点出现中断的地方，可以辅助维持光刻的均匀性。一些设计中 WL 条上存在间隙。这样的设计避免了大量前面所描述的缺点，特别是与多晶相对于扩散的位

置关系以及不对称放置接触孔情况下带来的问题。另外，对与多晶上衬线和反衬线相连的互连端头也进行了优化，使扩散区上的多晶尽可能是直线，从而避免对套刻未对准的敏感性。为了与采用 MOPC 修正过的单元进行对比，所绘的单元版图如图 11-14a 所示。

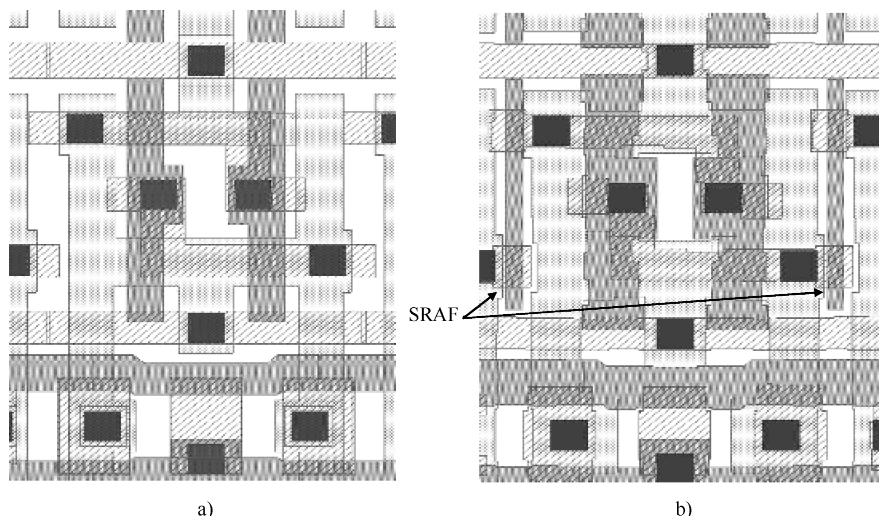


图 11-14　采用已得到良好表征的 MOPC 设计的位单元
a）采用 90nm 规则绘制的位单元　b）采用 MOPC 和 SRAF 的位单元

在纳米 CMOS 时代，位单元尺寸变得足够小，使得节点电容数值已降低到传统的快速提取方法本身的误差容限范围内[23]。如果使用传统方法通过快速提取得到的网表来建立一个仿真平台进行模拟分析，则各种误差叠加在一起。因此当加工好的硅芯片从制造厂返回到我们手上进行测试时，通常会惊奇的发现当初的建模是错误的。从 90nm 节点开始，为了实现位单元寄生的提取，开始使用场求解工具是明智的。

11.2.3　应对波动性的模拟电路策略

必须应对波动性对模拟设计者来说不是一个新概念。一方面，工艺尺寸缩小使一块芯片上数字电路的性能和功能得到指数性的改善和提高成为可能，但是另一方面，在许多前沿性方面，模拟设计者面临着更多的挑战。本章我们讨论了模拟设计者面临的波动性问题。虽然应对波动性问题不是新课题，但是有些问题变得更加严峻，同时又新出现了更多的问题。对这些效应进行精确的建模对模拟设计者来说非常重要，但是这也是一个主要的挑战。表 1-1 总结了可能影响模拟设计者的建模挑战。

许多模拟电路要求良好的器件匹配。我们将讨论一部分匹配问题和可以缓解或使它们对模拟电路影响最小的可能的方法。下面列举了影响匹配的主要来源。

与设计相关的来源：

（1）不对称性（导致对未对准的敏感性）；

（2）小几何尺寸（窄宽度效应；短沟道效应；较大的 V_{th}波动）；

（3）邻近效应［阱邻近、多晶邻近（线性邻近效应）、微负载刻蚀效应］；

（4）阱和地接触点的位置（体效应差异）；

（5）水平和垂直效应；

（6）温度差异；

（7）STI 应力效应；

（8）扩散和多晶图形光刻后直角弧化（在纳米 CMOS 电路中对设计有很强的影响）；

（9）镜像版图效应（电容、R_{sd}、未对准）。

与工艺、器件和电应力有关的来源：

（1）随机掺杂波动；

（2）穿过栅的杂质通道；

（3）多晶 L 波动、L_{eff}波动；

（4）由于天线效应引起的衰退；

（5）负偏压温度不稳定性（NBTI）；

（6）热载流子注入（HCI）；

（7）OPC（修正过度或者不足、多晶互连末端回缩、多晶缩颈和图形光刻后；直角弧化、扩散图形光刻后直角弧化）；

（8）金属密度波动［ILD 厚度波动（见图 11-15）、电容波动］；

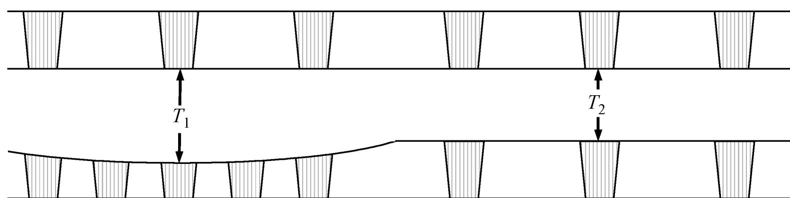


图 11-15 互连密度波动导致 ILD 厚度的波动

改善匹配的技术：

（1）增加输入信号摆幅。只要可能，增加输入信号的摆幅。例如，与需要检测低摆幅信号的读出放大器相比，读出放大器触发器对器件匹配更加不敏感。显然，我们不能增加读出放大器的输入摆幅，除非我们愿意牺牲速度或使用非常大的位单元。读出放大器必须依赖其他匹配技术。

（2）创造一个器件版图库。这可以确保使用尺寸相同的器件。如果需要一个较大尺寸的器件，可以将几个器件并列组合使用。图 11-16 所示的运算放大器的版图说明了这种并列组合器件的运用。

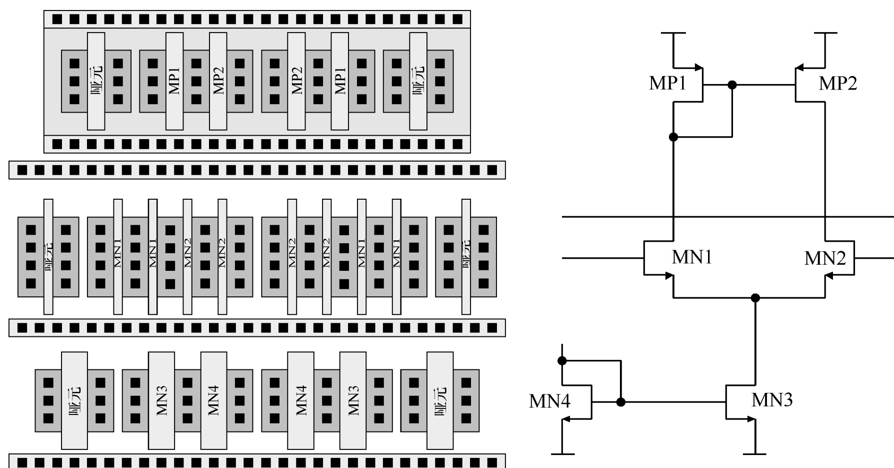


图 11-16　使用器件并列组合的版图

（3）使用哑元晶体管。这可以使由刻蚀微负载效应和多晶邻近注入散射引起的邻近效应的差异最小化。哑元晶体管的使用也可以减轻 STI 应力效应（见图 4-20c）。

（4）所有晶体管同向排列。如前所述，正交排列的晶体管会导致较高的 CD 波动和邻近效应波动。因此使所有的晶体管同向排列很重要，与其他关键电路（如时钟缓冲器、读出放大器和位单元）中的晶体管一起，很难使被减缓的 CD 波动达到最小。

应该避免的版图：

（1）避免镜像；而是采用步进结构，并且可能的话，采用共质心布局（见图 11-17）。一个镜像的晶体管版图与光刻套刻误差相结合，将会由于寄生电容和晶体管漏和源电阻的波动，导致失配。

（2）避免采用最小器件尺寸宽度或长度。因为杂质注入是一个统计事件，在较小区域上发生的较大注入波动，将使最小宽度器件呈现较大的阈值电压（V_{th}）波动。由于扩散图形光刻后直角弧化，作为“哑铃”形状版图的结果，采用最小宽度的晶体管也会发生形状失真（见图 11-18）。由于纳米 CMOS 时代短沟晶体管 V_{th}陡峭的滚降，采用最小沟道长度的晶体管会导致更大的 V_{th}变化。多晶 CD 的很小变化将会导致 V_{th}发生较大的变化（见图 11-19）。因此，采用最小沟道长度的晶体管的 V_{th}匹配性能很差。

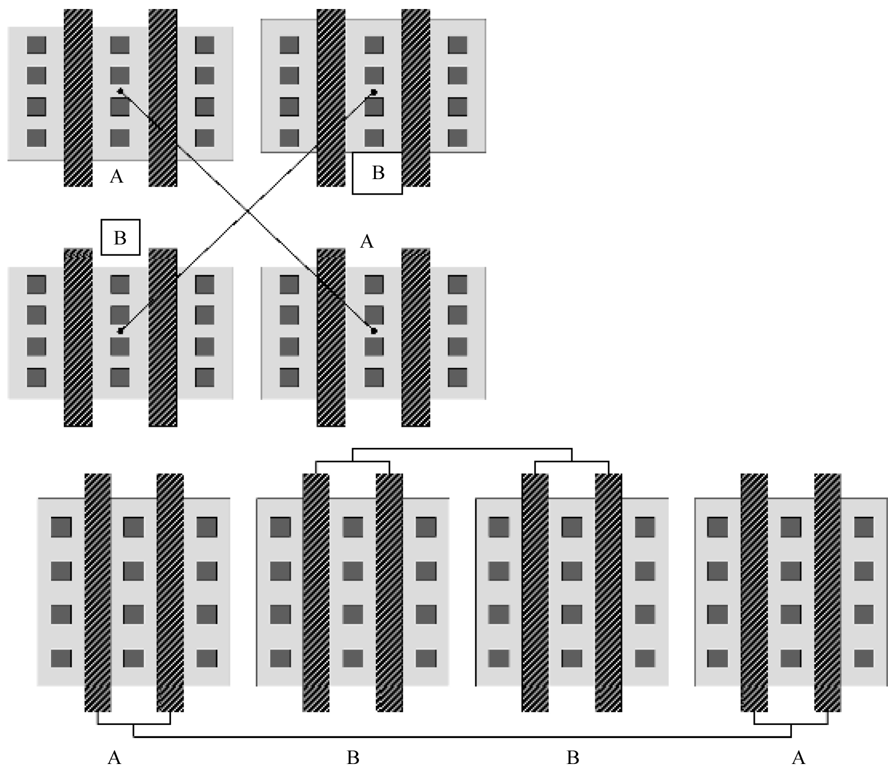


图 11-17　可以实现更好匹配的共质心版图；一些可能的布局

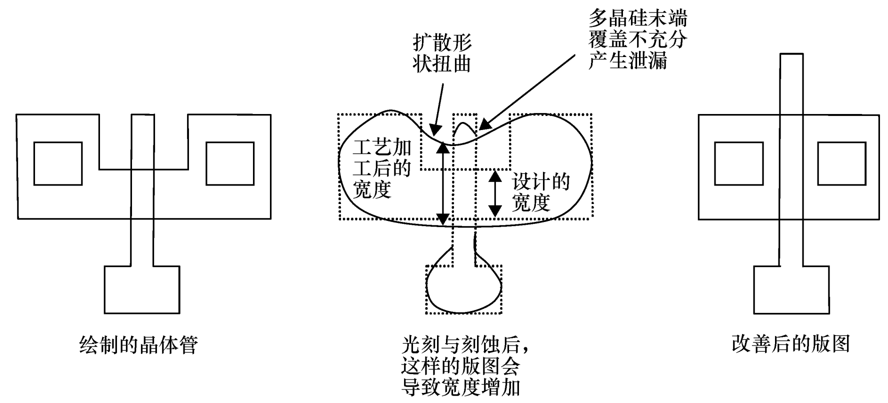


图 11-18　扩散图形光刻后直角弧化

（3）由当前的掩膜制造方式，栅弯曲的晶体管会导致尺寸的变化。对栅弯曲的晶体管，确定其宽度和长度也很困难，因此应该尽量避免在模拟设计中使用，特别是当匹配很重要时。栅弯曲晶体管对光刻套刻误差也极度敏感。

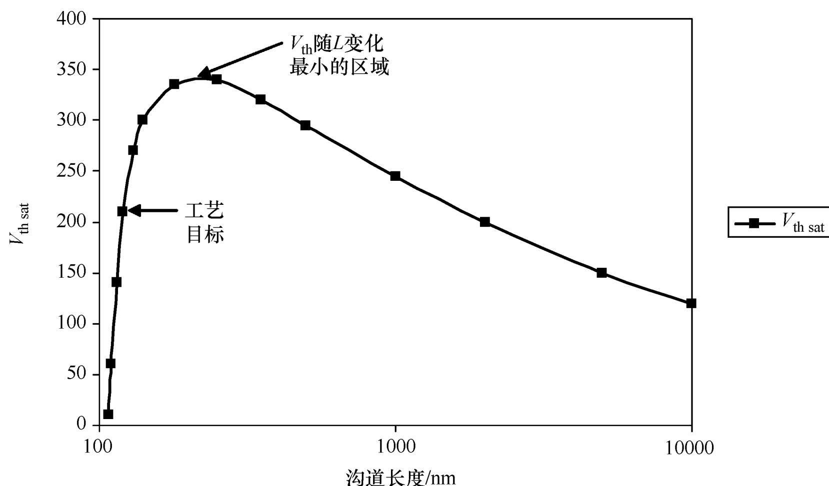


图 11-19　sub-100nm 晶体管典型的 V_{th}滚降特性

（4）应该避免在扩散边界或拐角处放置接触。因为拐角的光刻失真和光刻套刻误差会导致接触电阻的波动。

（5）在任何存在直流电流的连接中，特别是对差分对，应该禁止使用多晶跨接。较高的方阻（即使是硅化物多晶，其方阻也比金属高两个数量级相对的）会产生较大 *IR* 压降，导致失调（见图 11-20）。另外，对多晶和硅化物多晶来说，电阻的波动性也比金属的大。

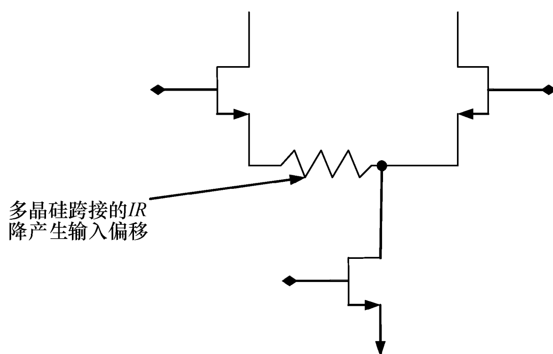


图 11-20　电流开关

版图的重点：好的版图实践

（1）尽可能在所有的模拟晶体管版图中，特别是对匹配晶体管，使用大于最小尺寸要求的扩散多晶末端交叠。有些生产线规定有特殊的设计规则，仅允许使用几种规定尺寸的交叠；版图中所有的实际交叠尺寸将被转化为与设计规则中允许尺寸最靠近的一个多晶末端交叠尺寸。在版图设计中必须考虑到这种设计规则的规定。否则，即使在版图设计中已经在扩散区使用了大于最小值的多晶覆盖，但是在制备的实际光刻版上并没有得到扩展交叠的好处。

（2）增加多晶图形中转角处与扩散边界之间的间距，避免由光刻套刻误差通过多晶图形光刻后直角弧化效应引起的尺寸波动（见图 11-30 和图 11-10b）。

（3）仅使用矩形或正方形扩散区特性而不使用“哑铃”形的扩散，在靠近

多晶边界的地方也不要使用90°的拐角（见图11-18和图11-31）。

（4）在那些匹配很重要的低摆幅电路使用共质心版图（参见图11-17，为了清楚起见，省略了围绕有源晶体管的哑元晶体管）。

（5）使用更宽的晶体管，并且将巨大的晶体管拆分，使得光刻套刻误差不会影响寄生电阻和电容的匹配（见图11-4）。较宽晶体管的使用还能缓解注入统计波动引起的器件参数波动。

（6）为了得到更好的输出阻抗（$I_{ds\ sat}$对 V_{ds}的变化率更低）和减小 CD 波动的影响，使用沟道长度更长的晶体管（约为最小规则的5倍）。

（7）使用完全接触的扩散，这将减小接触电阻波动。考虑到同样的原因，至少使用两个或更多（如果间距和版图允许）的通孔。

（8）使互连寄生匹配，包括 Miller 电容。还应该使输入以及输出互连长度匹配。每一层次使用的输入互连必须匹配，这样在等离子刻蚀期间，作为天线应力效应的结果，匹配晶体管会受到相同的应力。

缓解波动性影响的电路技术

（1）在余度的限制范围内保持电流源的栅驱动尽可能的高，使电源波动的影响和电流源栅上的噪声耦合最小。在源节点与电源或地相连接的电流源中，一种较好的处理方法是，将栅参考节点旁路到电源或地，可以改善电源和地的噪声抵抗能力。

（2）在电荷泄漏会引起失效的地方，使用厚氧化层晶体管作为电容。例如，PLL 中的环形滤波电容就应该使用厚氧化层晶体管，因为环形滤波器上的电荷决定了压控振荡器（VCO）的频率。环形滤波电容上的电荷泄漏会导致静态相位偏差，并且如果泄漏超过电荷泵的电流还会引起失效。在高倍数 PLL 中，环形滤波电容泄漏将引起 VCO 的频率漂移并且导致更高的抖动。例如，一个 ×20 的 PLL 每20个时钟周期才接收电荷泵的电荷。多路复用器的复用越高，在电荷泵更新期间，VCO 输出不接受电荷泵电荷的周期数就越大。对 PLL 环形滤波器，因为面积问题，并且容量不易控制，因此不推荐使用金属电容。当工艺升级到低 k 介质时，这将改变 PLL 的带宽。金属电容应该用于轨到轨之间需要线性电容的应用中。电荷泵晶体管亚阈区泄漏也会引起相似的问题。然而，栅泄漏随着温度增加的变化不大，而亚阈泄漏与温度的关系非常密切。

影响模拟电路性能的新物理效应

RSC 效应引起长沟道的 V_{th}低于短沟器件。如图11-19所示，当 L 增加时，V_{th}增加到峰值后接着下降。当沟道长度大于1μm 时，V_{th}将低于例子中工艺的工艺目标。为了实现最小的 V_{th}波动，必须将沟道长度设定在曲线的峰值处，采用这个沟道长度可以使 V_{th}波动最小。

作为 V_{th}变化的结果，另一个重要的效应是驱动电流在特定的偏置条件下随着温度增加，而在其他偏置条件下却随着温度下降。如图11-21所示，当晶体管

的 V_{gs}值小于约 630mV（低于交叉点）时，其驱动电流随着温度增加，但是当其 V_{gs}值大于约 630mV 时，驱动电流随着温度减小。这对运算放大器的开环增益有极大的影响。它也使得读出放大器的增益、电流开关或差分对、以及比较器随着偏置发生变化。这些电路总会有一部分晶体管被偏置到图 11-21 所示曲线的交叉点下方，而其他晶体管则被偏置到交叉点上方。

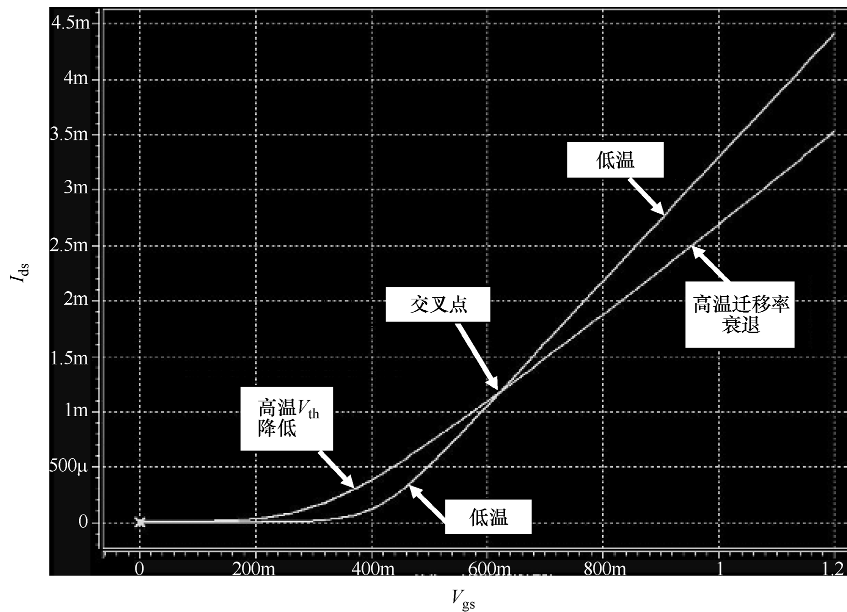


图 11-21　温度对 sub-100nm 晶体管 I_{ds}的影响

如图 11-22 所示，直到沟道长度为 10μm 范围，halo 注入晶体管的 DIBL 随着沟道长度持续下降（漏感应阈值电压漂移，DITS）。除非这个效应被很好地建模，

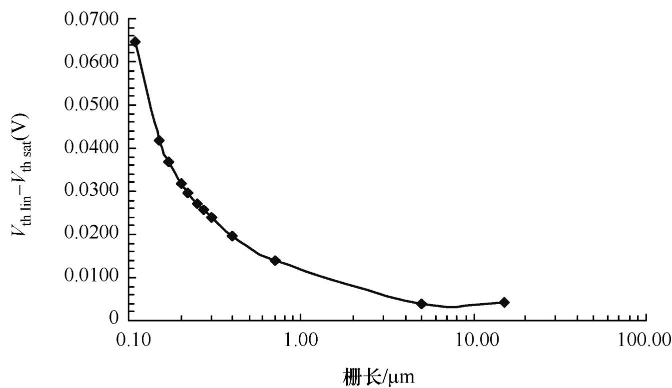


图 11-22　漏感应阈值电压漂移

否则在为了提升逻辑性能而进行工艺缩小时给器件产品特性带来出乎意料的变化。

图 11-23 所示为阱和多晶邻近对器件 V_{th} 的影响。考虑到这种带来新的负面影响的邻近效应，放置匹配晶体管时必须非常小心。

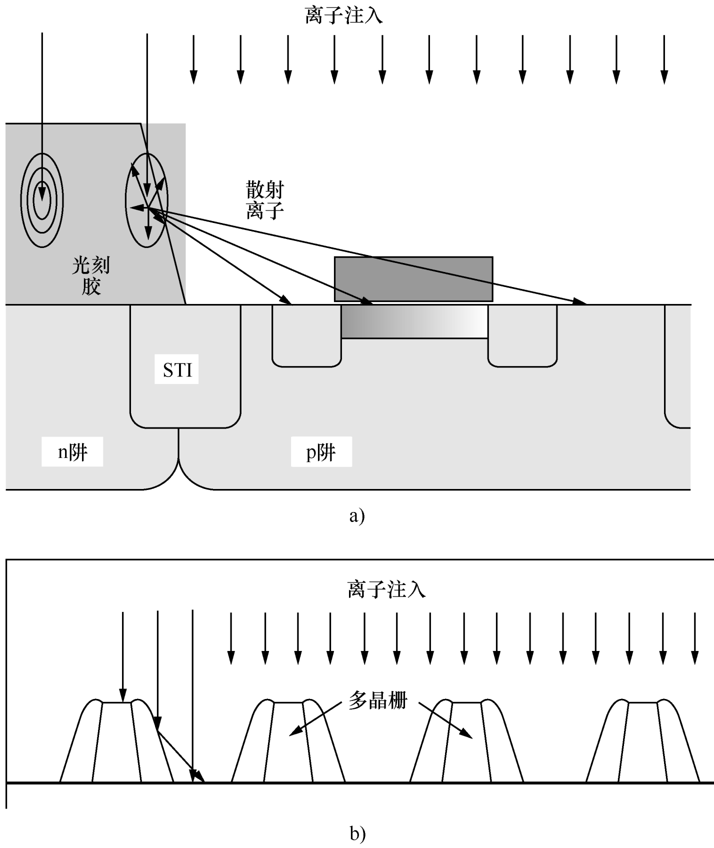


图 11-23 a）阱邻近效应 b）多晶邻近效应

电源纹波抑制比（PSRR）

逻辑晶体管的缩小已经导致了电源电压的减小，这限制了特定电路技术的使用，例如能提供更好 PSRR 的折叠式共源共栅电路。新出现的电路仍然可以提供适合电源余度的更好的 PSRR（细节请参看第 4 章）。其他的电源问题包括通过外延生长的电阻衬底产生的噪声耦合。在混合信号设计中，将最终需要采用三阱工艺将数字噪声与敏感的模拟电路隔开。为了减小容性耦合噪声，保持较小三阱的面积非常重要[27]。保护环的使用以及更策略的放置衬底和阱接触使衬底噪声变成共模噪声，这在差分电路中可以被排除。

电源阻抗的封装和系统建模非常重要，特别是对高性能芯片设计。在这样的

设计中，$L(\mathrm{d}i/\mathrm{d}t)$是最重要的电压降，并且电源阻抗必须被设计为可以将$L(\mathrm{d}i/\mathrm{d}t)$与$IR$降之和保持在设计预算内。对高性能处理器来说，这个预算通常是电源电压的 10%。这将需要使用片上解耦电容、封装电容和系统板上电容。最小值的片上耦合电容应该是芯片等效开关电容 C_{eqv}的 10 倍，其中：

$$C_{\mathrm{eqv}} = \frac{\text{芯片功率}}{V_{\mathrm{dd}}^2 \times \text{频率}}$$

芯片功耗使用最坏情况矢量决定，这意味着 C_{eqv}值将更大并且可以更好地维持峰值功率需求期间的芯片功耗。因此，由于电源具有良好的解耦，峰值功率需求期间，由电源电压降引起的性能衰退将被保持在最小值。为了满足高性能系统和电源阻抗的目标，封装和系统板级建模也是设计中非常重要的部分，而这超出了本书的范围。

在采用外延工艺的混合信号设计中，共用 V_{SS}但是分别使用不同的 V_{dd}电源将在数字部分形成最低的噪声耦合[12]。为了给一些设计提供必须的余度，就可以使用高于数字电路的电源电压。共用 V_{SS}也简化了静电放电（ESD）保护（在纳米 CMOS 时代，关于 ESD 保护的详细讨论参看第 5 章）。

11.2.4　应对波动的数字电路策略

数字电路一般对工艺波动有更好的适应性。然而，有一些数字电路，包括自计时电路和匹配时延的电路，都对工艺波动极为敏感。自计时主要用于像缓存这样的嵌入式存储器中。在时钟频率较低时，其应用非常广泛。为了减小存储器的存取时间，自计时技术用于产生边沿，从而为读出放大器（Sense Amplifier，SA）提供（内部）时钟，这样在时钟周期的开始阶段，就可以准备好存储器数据，从而能够实现一个周期完成读取，包括对存储器数据的逻辑操作，进而实现更好的性能。随着（内部）时钟频率的升高，嵌入式 SRAM 的存取时间小于时钟周期，这样可以有大量的边沿用来对 SA 实现钟控。因此，在这种情况下，不需要强制自计时产生大量边沿。其他唯一需要自计时的情况是，通过地址改变之前不对 SA 提供时钟的方式节省功耗，而时钟设计也要求通过钟控门以减小时钟功耗。对 SA 进行钟控仍然比自计时更简单且更具鲁棒性。人们设计了许多的方案，可以在必须采用自计时的情况下，能够缓解波动对设计鲁棒性的影响。下面讨论一种用于 SRAM 的自计时方案。

自计时策略　传统的自计时存储器依赖于单个 SRAM 单元来驱动伪位线，然后转变为完全的 CMOS 电平及扇出以驱动 SA 的时钟线[28]。如图 11-24 所示，单个单元的自计时方法对工艺波动非常敏感，这些工艺波动引起单元驱动波动，导致更高的自计时时延波动。为了避免由于更高的自计时路径的时延波动引起的失效，就需要更大的裕量——这需要以性能为代价。

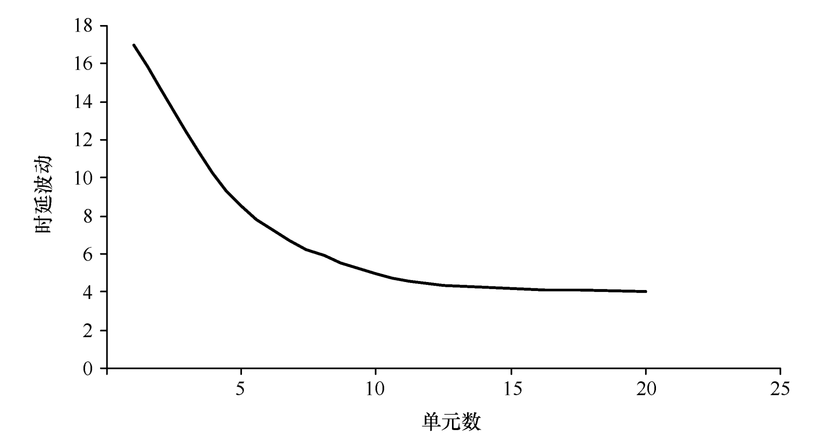


图 11-24　单元数量对自计时时延波动的影响

如图 11-24 所示，如果使用不止一个单元，可以实现较低的自计时路径时延波动[13]。使用几个单元可以使单元电流波动得到平均。图 11-25 是一个多单元自计时方案。为了使单元驱动波动最小，在阵列的边界、伪自计时列旁用另一列填充单元作为伪自计时列是很重要的。在存储器列中，金属密度非常一致，使控制 ILD 波动变得容易。由于金属互连的规律性，由化学机械抛光（CMP）引起的电阻波动保持为最小，如在存储器阵列中看到的，制造商为了金属密度而优化 CMP。大多数制造商都认识到必须减小存储器阵列电阻的波动，并且会优化存储器阵列的工艺。即使如此，仍然存在由于屏蔽层金属厚度和互连宽度波动而引起的电阻波动。

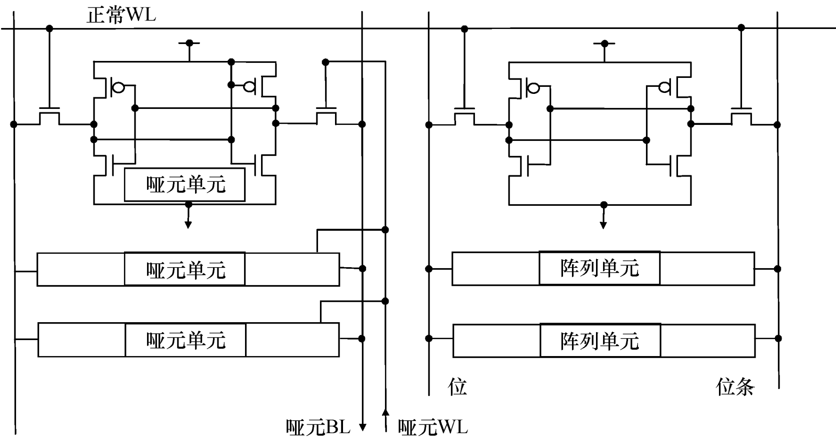


图 11-25　多单元自计时方案

自计时裕量　图 11-26 所示为一个典型的自计时设计中的竞态条件。图中 Out1 上的时延必须小于 Out2 上的时延；否则，将会发生功能失效。由于工艺、

电压和温度（PVT）的波动以及版图的差异，在实际生产出的电路中，由于一些局部效应在设计阶段并未被完全正确地建模或预见，可能会导致时延 2 小于时延 1。当自计时设计中出现这种情况时，会导致在一些频率下不能工作，包括非常低的频率，这时会要求设计返工以恢复基本功能。这是非常严重的并且代价昂贵的设计缺陷。为了预防这种情况，我们增加了仿真模型的余量以覆盖一些未能预见的影响，以减小这种功能失效的概率。

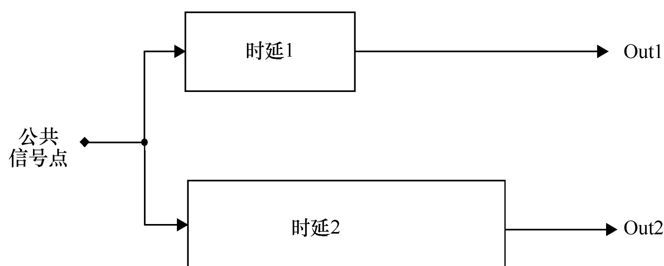


图 11-26　设计自计时路径

如前面描述的，由于一些未知的影响或电路没有得到充分优化，时延 2 的速度可能与时延 1 不匹配。下面的分析将“余量”转化为那些能被用来验证自计时电路余量的具有物理意义的参数。图 11-26 所示的处于失效边缘的自计时电路可以表示为：时延 2 × (1 − M) = 时延 1 × (1 + M) 其中 M 是自计时裕量。将其简化，可以得到：

$$M \times (\text{时延 }1 + \text{时延 }2) = \text{时延 }2 - \text{时延 }1$$

因此有：

$$M = \frac{\text{时延 }2 - \text{时延 }1}{\text{时延 }1 + \text{时延 }2}$$

一般在所有实际工艺角下进行仿真时，将预布局布线时的 M 设定为 0.25，而将版图后仿真时的 M 设定为 0.15。对更实际工艺角覆盖，推荐使用统计模型。关于统计模型的详细介绍见 11-3 节。对于给定的自计时裕量，每个自计时路径必须采用金属编程选项，以保证在所有实际工艺角下使裕量至少再增加 30%。如前面提到的，自计时竞态失效对芯片来说是灾难性的；金属编程选项的增加会导致快速的环锁定。金属选项必须被设计为可以影响一层并不超过两层中的裕量波动。这非常重要，因为掩膜的成本在不断增加，特别是对于纳米 CMOS 工艺节点。如果可能，在尽可能高的金属层中设计编程方式，可以在需要改变自计时裕量的情况下，加快产品的制造周转时间。

慢节点导致的时延波动　慢节点包括高扇出节点、不重复的长互连以及通过传输门和级联传输门的信号。正像不重复的长互连那样，传输门对信号来说相当

于是一个大的电阻。如果一条信号路径上超过两个传输门（未采取缓存），就相当于是一个必须被处理的非常慢的节点。就像信号穿过级联传输门和不重复的长信号线的情况那样，慢节点也可以是驱动很弱的节点。被弱信号驱动的节点易受接收器所在的远端节点中的噪声耦合的影响。还有另一个障碍影响所有慢节点，包括高扇出节点。如图 11-27 所示，由于慢节点上输入信号的缓慢上升，接收机输入跳变点的波动被转换为较大的输入时延波动。由于工艺加工会引起 P 管 N 管之间的差异，进而会影响门输入阈值或跳变点，维持输入边缘的转换斜率则可以使设计能够更好地容忍 P 管与 N 管之间的差异。

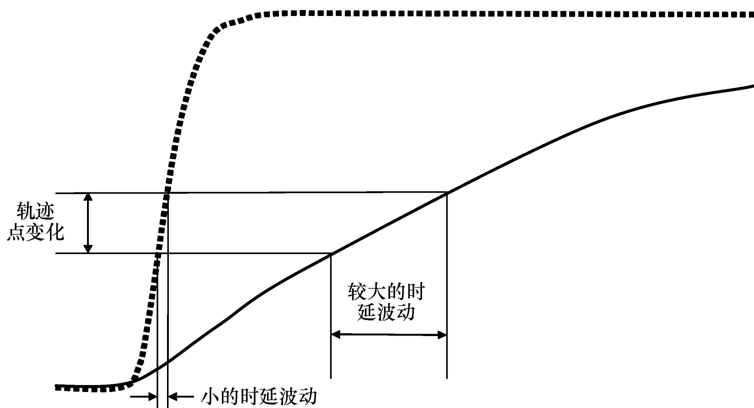


图 11-27　跳变点变化与时延波动比对

在一些电路中，例如算术模块，如果使用传输门加法器，在数据路径上就会有传输门。在某些情况下，数据路径上会有几个传输门串联，除非设计者在级联的全加器之间增加缓存器。这将增加了关键路径上的时延。通过改用差分级联电压开关（DCVS）逻辑可以缓解这个问题[31,32]。

脉冲寄存器时钟发生器设计策略

跳变点匹配：脉冲寄存器工作过程和设计不是本书要介绍的内容；对其详细的讨论请参看其他电路设计文章。但是，想要能够理解下面关于工艺波动对脉冲发生器和脉寄存器工作过程的影响的讨论，则需要对脉冲寄存器工作过程有全面理解。图 11-28 中是一个用于脉冲寄存器的脉冲发生器典型电路。Inv1 到 Inv3 形成了一个确定脉冲发生器脉宽的时延链。脉冲发生器的脉宽波动对脉冲寄存器的保持时间有严重的影响。Nand1 和 Inv1 的输入跳变点必须匹配；否则，脉冲宽度会随着全局时钟边缘斜率的波动而改变。更长的脉冲输出宽度将会要求更长的保持时间，同时导致更长的透明时间。如果输入到脉冲寄存器中的短脉冲没有及时得到适当的平衡，即使对这个短脉冲存在最大时间路径，由于更宽的脉冲宽度引起的更长透明期也会引起保持时间的问题。

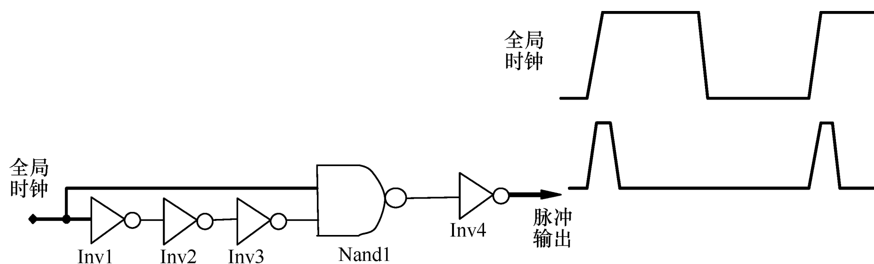


图 11-28　典型的脉冲寄存时钟发生器

我们从图 11-28 所示的 Inv1 开始，考虑 Nand1 的跳变点比反相器链更高的情况。当全局时钟上升时，Inv1 首先动作并启动时延链，而 Nand1 尚未对全局时钟输入做出反应。这将实际上缩短了脉冲发生器输出脉冲的宽度，因为反相器时延链跳变更早地触发了脉冲发生器输出的上升沿。从 Inv1 开始动作到 Nand1 触发之间的时延等于脉冲发生器输出脉冲宽度减小的量。如图 11-27 所示，时钟上升时间波动会改变这个时延，从而改变脉冲宽度。有多种原因会导致时钟上升时间的变化，而该变化会影响芯片的保持时间并引起灾难性的失效。Nand1 跳变点低于 Inv1 的触发条件会增加脉冲宽度和脉冲寄存器要求的保持时间。在基于单元的设计中，脉冲寄存器表征条件是假定 Inv1 和 Nand1 的跳变点完全匹配。如果 Inv1 和 Nand1 的跳变点不匹配，将引起脉冲寄存器实际要求的保持时间改变，从而可能产生保持时间失效。

设定输入跳变点稍稍低于 $V_{dd}/2$（低于中点的 1/3），但不要太低；否则，会受到地电位反弹影响。这是因为在时钟上升沿的较低位置，边缘误差较低。由于脉冲发生器仅参考时钟上升沿，因此这一技术可以确保实现更精确的时钟参考和更低的时钟边沿时延。

脉冲发生器输出波形峰值：脉冲宽度必须足够宽，以确保在各种的实际情况下，对所有负载条件，脉冲都能到达 V_{dd}。这保证了脉冲宽度是确定的。如果在所有的负载条件下脉冲宽度都能够到达 V_{dd}，脉冲将总是在相同的 PVT 条件下从相同的电压释放电荷，因此脉冲宽度将是确定的。这就消除了脉冲宽度的波动（不包括 PVT 条件引起的波动）。使时钟脉冲达到 V_{dd} 的其他原因是确保总是在寄存器的时钟输入端有相同驱动电平，从而避免由于门驱动的波动导致的建立时间和保持时间的波动。

脉冲发生器时延与数据路径时延密切相关：为了保持低功耗，由 Inv1 和 Inv2 形成的时延链必须采用最小尺寸晶体管构成。这里我们不得不以功耗为代价来减缓工艺对数据时延的影响。所选器件必须足够大，从而使得时延不会主要取决于寄生参数。就像以提高速度为目标而优化的数据路径上那样，沿着时延链的寄生必须被

最小化。在各种实际可能发生的极端情况下，时延链提速的速率必须与数据路径的相匹配，避免出现保持时间冲突。如果数据路径提速快于时延链，特别是对动态脉冲寄存器，就能够避免动态寄存器的输入数据在脉冲复位前发生变化。

时延链上的最后一个单元（Inv3）必须与脉冲发生器驱动的逻辑寄存器有相同的堆栈高度。脉冲发生器产生时钟脉冲，如果接收时钟脉冲的寄存器不是简单的寄存器而是动态逻辑寄存器，时延链上的 Inv3 必须与寄存器前面的动态逻辑有同样的堆栈高度（参见图 11-29）。这使得时延链可以在各个工艺角上都能够满足逻辑时延。图 11-30 与图 11-31 是两个实例，说明需要放宽间距规则以及多晶端头覆盖，以减小由于光刻引起的多边形图形失真导致的器件波动。

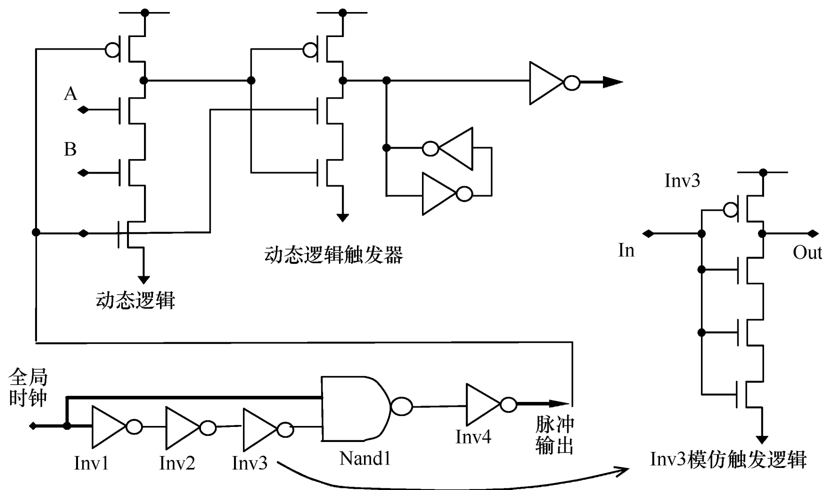


图 11-29　脉冲发生器的时延追踪技术

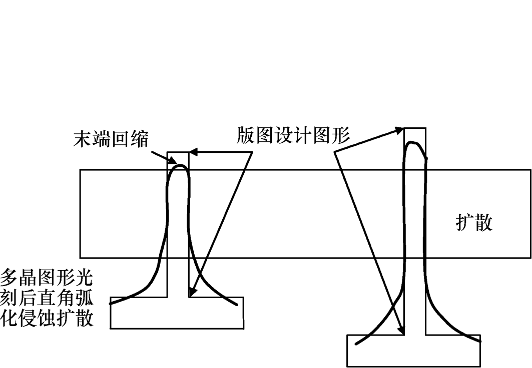


图 11-30　光刻引起的多晶图形直角弧化

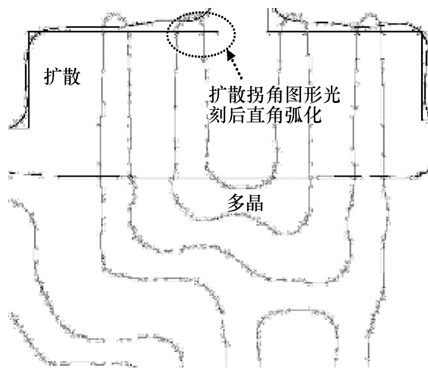


图 11-31　扩散图形拐角处的多晶端头覆盖不良

11.3　纳米 CMOS 工艺角建模方法

SPICE 建模已经成为使设计者可以确定必要的设计裕量以满足现代 IC 电路严苛要求的最关键部分。随着不断增加的速度要求，裕量在持续减小，这迫使设计者更加依赖模型以精确反映工艺状态，包括预期的分散性。传统的模型开发方法一直是建立与受控生产线工艺标称值一致的标称值模型，然后再针对数字逻辑中的最坏情况开发拐角模型。表征工艺分散性的方差并不随着特征尺寸的缩小而同等的缩小，由此引入的误差越来越大，特别是对深亚微米工艺。为了更精确地表征工艺，就对统计模型有一个实际需求。图 11-32 所示为不同级别的工艺波动。工艺流程中的每个级别都会给器件性能增加附加的分散。理解每级的影响对开发精确统计模型非常重要。

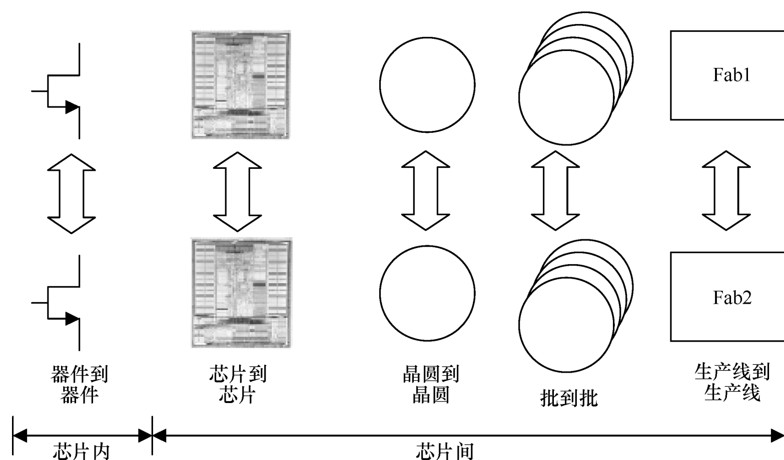


图 11-32　各种级别的工艺波动

11.3.1　统计模型的需求

采用工艺角模型方法考虑的工艺组合与实际情况不相符合，会导致“过于保守的设计”，特别是当设计裕量变得更小时。这一点可以参见图 11-33，图 11-33所示为从多个晶园批得到的 NMOS 和 PMOS 的 $I_{D\ sat}$测量数据的散点图。其中，快——慢和慢——快（FS 和 SF）拐点很少出现。这从工艺的角度看是可以理解的，因为 PMOS 与 NMOS 器件仅局部相关。例如，如果我们考虑具有波动性的各种参数，像氧化层厚度、栅长度、栅宽度、沟道掺杂和 halo 注入等，它们中间一些参数（如氧化层厚度和沟道长度）的波动对 PMOS 和 NMOS 器件是相似的，而其他参数并不相关并且会独立的波动。另外，工艺的波动包括局部范围波动和

全局范围波动两个成分，而工艺角并不能区分这两个分量，因此基于拐角模型不可能确定器件之间局部范围波动性的影响。

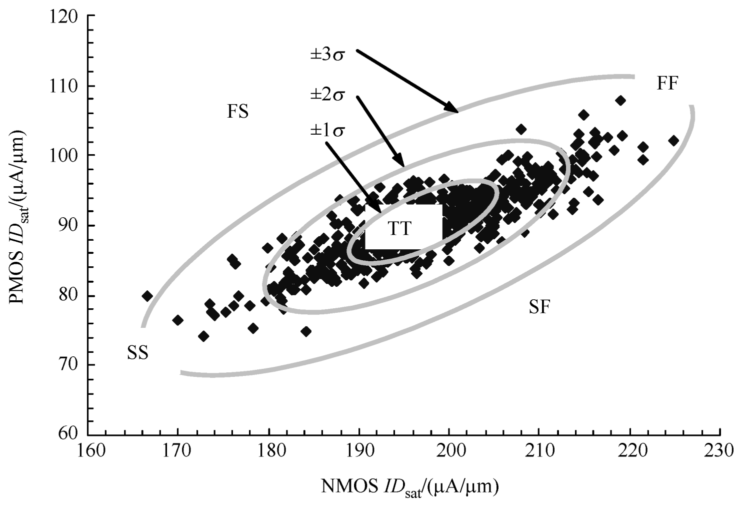


图 11-33　PMOS 和 NMOS 器件的工艺波动图

对模拟电路而言，很难确定最坏情况的拐角。快/慢的概念不一定适用。对运算放大器，高增益/低增益可能很有意义，但是很难确定哪个数字工艺拐角对应于放大器的高增益情况，因为它依赖于放大器的结构特点。当子模块被组合成像数据转换器那样更复杂的系统时，确定哪个工艺拐角代表最坏拐角变得更加困难。如果模拟电路仿真中，特别是已经给定了有限的模拟电路设计空间的情况下，采用数字工艺拐角将导致模拟电路的“过设计”。电路的过设计会导致复杂度增加、晶片尺寸更大、并且可能会失去市场窗口，因此如果有可能的话应该尽量避免。如果我们考虑工艺中几种可变参数的波动性，总的方差可以表示为

$$\sigma_{\text{total}} = \sqrt{\sigma_{t_{\text{ox}}}^2 + \sigma_{\text{L}}^2 + \sigma_{\text{W}}^2 + \sigma_{N_{\text{p}}}^2 + \sigma_{N_{\text{n}}}^2 + \sigma_{\mu_{\text{p}}}^2 + \sigma_{\mu_{\text{n}}}^2 + \cdots} > > 3\sigma$$

如果必须满足这些极端情况下的性能要求，采用上式表示的总方差会导致电路严重“过设计”。

采用统计建模可以使设计者在制造前就能估计给定设计的功能成品率。这个信息对于在设计阶段而不是在制造后进行的参数权衡设计非常重要。设计者可以分析设计中的子模块以确定每个部分对整个系统成品率的贡献，从而使设计重点可以被放在设计最关键的部分。设计者也可以评估器件尺寸对功能性成品率的影响。

11.3.2　统计模型的使用

统计模型是基于实践，测量波动源并且将这些波动转化为 SPICE 模型参数

的波动。第一步是识别独立的波动因素并确定它们的长期波动。图 11-34 是一个具体实例，图中显示了一段时期以氧化层厚度表示的电容等效厚度的波动。这些数据信息被转化为直方图，可以进一步提取得到表征波动特征的均值和标准偏差值。这些特征值然后被结合进模型，将独立的模型参数建模为其标称值加标准偏差变量。可以按照这种方法处理的物理参数包括掺杂浓度、氧化层厚度、迁移率、栅宽和栅长等。重要的是这些选择的参数可以被正确的加到 SPICE 模型中以确保它们的影响可以被正确的仿真。例如，普遍使用的简单办法是改变器件的阈值电压以观察工艺波动的影响，但是这并不能正确的获得背栅偏置的情况，因而会产生不正确的结果。因为 SPICE 模型并非都具有完整的物理含义，这使得这个任务变得很困难。

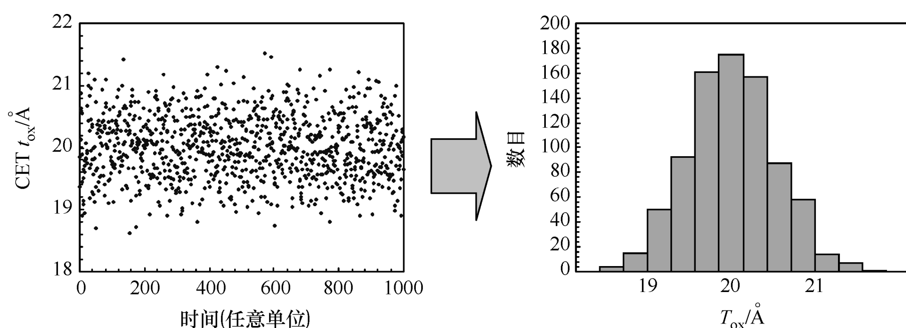


图 11-34　给定工艺的氧化层厚度随时间的波动

下一步是模型关联。通常像阈值电压 V_{th0} 这样的参数被设定为固定值，例如取 $V_{th0}=0.4$，考虑到波动的影响，现在就成为 $V_{th0}=0.4+\text{VTH_PVAR}$，其中 VTH_PVAR 定义为 AGAUSS（$\overline{M}$，$\sigma$，$N$），这里 $\overline{M}$ 代表均值，σ 是标准偏差，N 代表 N 倍标准偏差 σ。采用这种方式并不能获得阈值电压对氧化层厚度的依赖，因此最好表示为[36]：

$$V_{th0}=V_{FB}+2|\phi_F|+\frac{qN_S x_{t1}+qN_P(X_{dep}-x_{t1})}{C_{OX}}$$

式中，$C_{OX}=\varepsilon_{OX}/t_{OX}$，$t_{OX}=\overline{t_{OX}}+\sigma_{t_{OX}}$，$N_S=\overline{N_S}+\sigma_{N_S}$，$N_P=\overline{N_P}+\sigma_{N_P}$，$V_{FB}=\overline{V_{FB}}+\sigma_{FB}$。相关参数的含义为：$N_S$ 是 $0\sim x_{t1}$ 之间的掺杂浓度，而 N_P 是 x_{t1} 与耗尽层深度 x_{dep} 之间的掺杂浓度。其他符号的含义与普遍采用的含义相同。使用这种阈值电压的表达式可以同时考虑像平带电压和沟道掺杂这样的多个工艺参数。这也可以获得衬底偏置的影响，使整个仿真更加精确。一旦获得合适的参数值，进行多次仿真就可以获得能够在晶圆上测量得到的像阈值电压和 $I_{D\,sat}$ 这样的参数的分布。对比实际的测量结果与仿真得到的分布，就可以验证模型产生的分布的

正确性。

对两种类型的器件，每个参数的标准偏差一般不同。同样每个参数的标准偏差与器件尺寸也有很强的相关性，对沟道长度依赖性更强，特别是在沟道长度很小的情况更加明显。图 11-35 所示为深亚微米工艺下两个并排放置以提供最大程度匹配的相同 NMOS 器件之间阈值电压差别（这种差别只反映局部范围内参数的分散性）随器件尺寸波动的情况。这些数据没有包括能进一步增加参数分散性的器件不同放置位置的影响。局部范围内参数的分散性可能对数字电路来说不太重要，因为数字电路特性主要取决于参数平均结果，特别是对较深的逻辑级别。但是对模拟设计它变得很重要。这种局部范围内参数分散性可以被用来确定像差分放大器这样的关键单元中器件的最优尺寸。能反映参数分散性的合适模型应该同时考虑局部范围（晶片内）和全局范围（晶片间）的参数分散性。工艺分散性可以用下式表示[34]：

$$\sigma^2(\Delta P) = \frac{A_{\Delta P}^2}{WL} + S_{\Delta P}^2 D^2$$

式中，$\sigma(\Delta P)$为工艺参数 ΔP 的标准偏差；W 和 L 分别为器件沟道宽度和长度。器件之间放置位置差别用 D 表示，而参数 $A_{\Delta P}$和 $S_{\Delta P}$是必须由测量确定的与工艺相关的常数。上式右边第一项代表局部范围参数的分散性，而第二项则代表了与器件放置位置密切相关的全局范围参数分散性。某些情况下，这个模型可能不能提供对工艺波动必要的洞察[35]。因此，最好是采用更多的分量构成方差，从而可以对可能引起分散性的各种位置进行深入的分析进而得到整个影响结果。可以进一步了解图 11-32 的细节，其中对每一级别均赋予一个方差分量。采用这种方式可以更深入的了解对产品成品率的影响，但是要针对各个级别获得关于分散性的有意义的信息会变得困难。

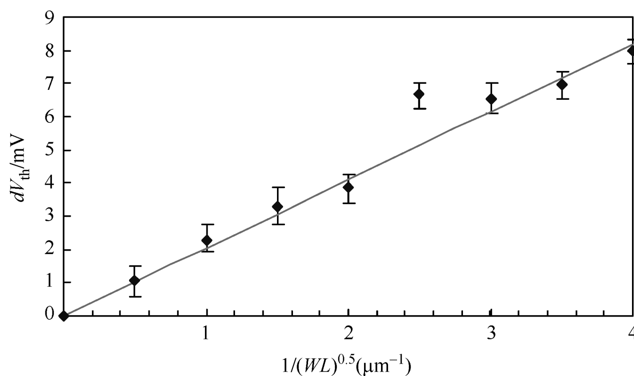


图 11-35　作为器件尺寸函数的阈值电压的波动

这种方法已用于估计锁相环的电荷泵中可以预期的电流失配程度。仿真结果如图 11-36 所示。这里假定设计可以处理 ±6% 的电流失配，而这些电流失配会导致 15 片芯片超出规范范围，或成品率约为 97%。如果认为这种成品率是足够的，那么就无需进一步改进设计。如果要求更高的成品率，电路就必须重新设计。这种重新设计可能要采用完全不同的电荷泵结构，或只要简单的重新确定关键器件的尺寸以降低参数分散性的影响。图 11-37 说明了阈值电压的分散程度是如何随着器件尺寸的增加而减少。图中 y 轴表示阈值电压的偏移，而 x 轴是归一化的器件尺寸（面积）。归一化是指对 100nm 工艺中的最小尺寸的器件进行归一化处理。通过有选择地放大关键器件的尺寸来减小整个系统的参数分散性是可能的。

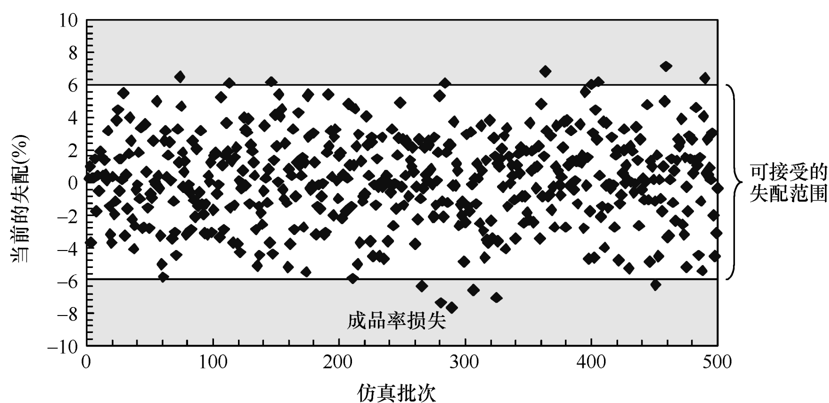


图 11-36　由局部范围和全局范围阈值电压分散性导致的电荷泵电流失配

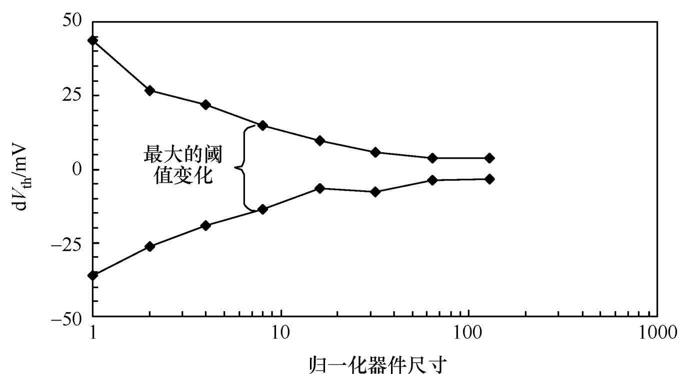


图 11-37　作为器件尺寸函数的阈值电压的分散性

11.4 BSIM4 模型的新特点

对深亚微米工艺，采用 BSIM4 模型使仿真的精度获得了很大的改善。BSIM4 模型包含了几个以前 BSIM3 模型忽略的重要特性，包括 halo 注入建模、栅感应漏泄漏（GIDL）、栅直接隧穿、和沟槽隔离应力效应。其中沟槽隔离应力效应已经在第 4 章详细讨论过了。

11.4.1 halo/packet 注入

采用 halo/pocket 注入的作用是减小沟道非常短的器件中出现的阈值电压滚降，但是对较长沟道的器件，这种注入会导致明显的 DITS。halo/pocket 注入增加了长沟器件的 g_{ds} 值，而对作为较长沟道器件主要应用领域的模拟应用，这是不希望的。图 11-38a 所示为 halo/pocket 注入的位置，而图 11-38b 所示为 100nm 工艺中产生的 DITS 效应。BSIM3 版不能完全描述这种输出阻抗的退化，因为

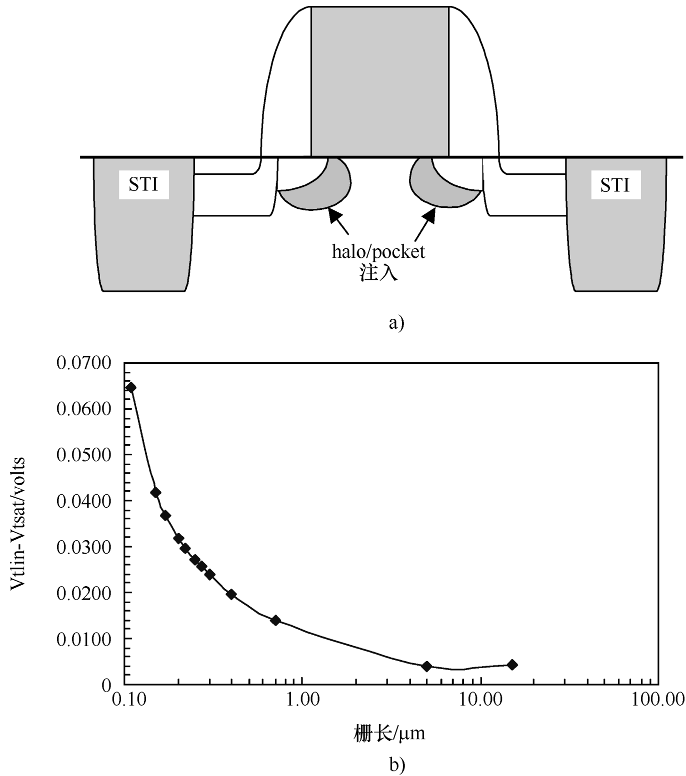


图 11-38 a）深亚微米工艺中使用的 halo/pocket 注入 b）100nm 工艺 DITS 模拟结果

BSIM3 版本中 DITS 未考虑 halo/pocket 注入的作用。不再将衬底视为均匀掺杂，模型中就能考虑 halo/pocket 注入的作用。因为 DITS 输出阻抗模型没有考虑体偏置效应，所以仍然存在限制。

11.4.2　栅感应漏极泄漏和栅直接隧穿

截止状态下泄漏电流中的各分量如图 11-39 所示，图中还给出了几代工艺中各分量影响程度的相对大小。在 90nm 和更小技术节点中，栅泄漏变得突出而成为越来越重要的因素，但是源漏泄漏仍然是最主要的问题。BSIM4 包含了栅泄漏模型。由于栅泄漏与栅上的电势有关，因此必须对每个栅偏置点的栅泄漏进行评估，这就导致需要更长的仿真时间。图 11-40 所示为不同栅长情况下总的归一化栅泄漏电流与器件宽度的关系，其中栅长范围为 0.2 ~ 15μm。图 11-41 所示为 100nm 工艺中薄栅器件的 GIDL 效应。GIDL 电流是毫微安量级。可以观察到其与体偏置之间存在微弱的依赖关系。

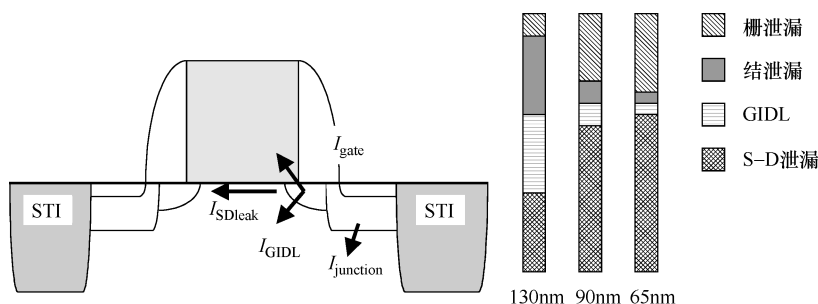


图 11-39　晶体管截止状态泄漏电流组成及不同分量相对大小随工艺的波动

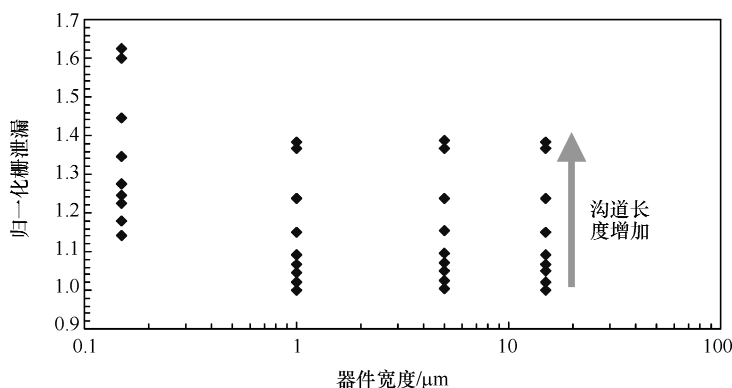


图 11-40　作为器件长度和宽度函数的总归一化栅泄漏

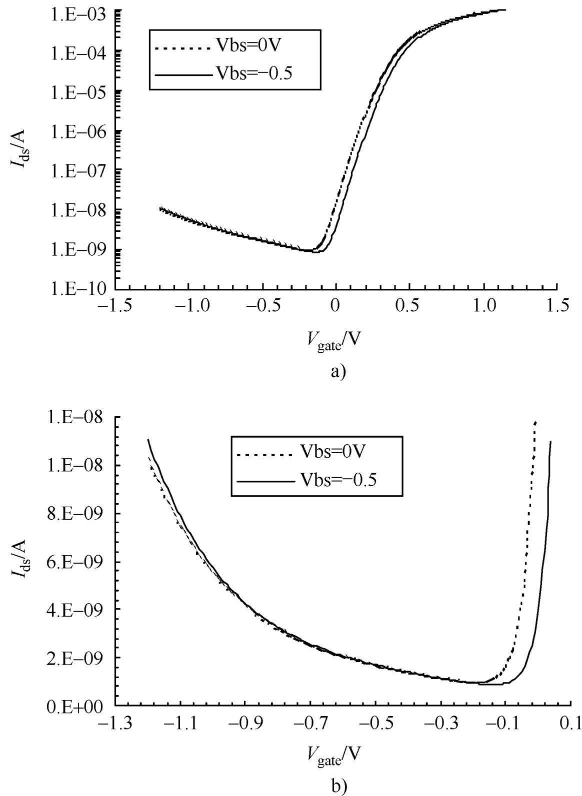


图 11-41 a）在较大范围内 GIDL 的模拟结果 b）局部范围模拟结果的放大，显示衬底偏置的影响

11.4.3 建模的挑战

虽然 BSIM4 是在 BSIM3 模型基础上有了重大改善，但是它仍然没有能考虑对器件性能有重大影响的所有因素。许多这些因素关系到如何放置器件以及相邻器件的物理位置：

（1）实现窄宽度的“哑铃”形状的器件；

（2）阱邻近效应；

（3）浅槽隔离应力效应（可以在版图设计后建立这些效应的模型）。

建议使用的方式是只要可能就应该避免出现会加剧这些效应的版图设计，因为它们很难被建模。这种方法会导致对物理实现的严重约束，增加整个芯片的尺寸。第二种方法是开发能考虑到这些效应的宏模型。应该对一个设计中最关键电路（例如 SRAM 单元）生成这些模型，以确保获得最高级别的精确度。这些宏模型应该采用参数化表示方式，以达到最大的灵活性。需要确定模型和先前测试芯片结果之间的相关关系，以确保模型的精确度。

11.4.4 与建模相关的问题

BSIM4 模型使用了非物理参数来实现短/窄器件的高精确度。非物理参数的使用使模型参数的提取过程更加复杂，因为短沟和长沟器件之间的关系比较复杂。如果模型不能充分描述这些物理效应，例如使用 Halo/pocket 注入技术时与掺杂密切相关的迁移率模型，如果模型不够精细，将导致模型所描述的器件与实际器件之间的一些差异。为了进一步改善模型的精确度，还需要建立反向短沟道效应（RSCE）模型。BSIM 模型的每一次进步带来了参数数量的大量增加，导致仿真时间和存储器要求的增加。这样，一个关键问题是平衡模型参数的数量和合理的仿真时间。

11.4.5 模型总结

BSIM4 模型极大地改善了 Halo/pocket 注入器件的建模，并且包含了在 90nm 和更小工艺电路设计中大量需求的栅直接隧穿模型。参数提取方式变得更加复杂，并且参数的数量也急剧增加。宏模型可以被用来对与版图相关的一些特定问题进行建模，但是必须建立其结果与实际硅器件测量结果之间的相关关系，以保证它们的精确性。迄今仍然存在几个应该被结合进模型中的很重要的效应，但是建立的模型不应该过于复杂，也不应该明显增加仿真运行时间。

11.5 总结

本章提出的一些原则可以被应用到许多其他电路和版图类型中以使波动对电路功能和可制造性的影响最小。当我们成功地将工艺尺寸缩小进入纳米 CMOS 时，应对这些波动问题将是所有设计方法（包括 ASIC 设计）的一部分或几部分。一些设计对波动比较敏感，应该在设计阶段特别小心，以预见可能的陷阱，从而可以采取针对性的设计策略和特殊预防措施，使波动不致对电路功能和可制造性产生负面影响。如果设计者希望得到满足设计目标要求并且成品率高的产品，必须学会创造对波动不敏感的电路。传统的波动概念已经从数字电路的工艺角方法进化到了要同时考虑芯片内和芯片间基本物理参数的统计波动。第 10 章我们更着重于关注电路设计中的可制造性设计问题，其中多数情况下也有助于减小由于波动性带来的影响。

参考文献

[1] International Technology Roadmap for Semiconductors, *http://public.itrs.net*.
[2] K. Bernstein, Design, process, and environmental contributors to CMOS delay variation, tutorial, *IEEE International Solid-State Circuits Conference*, Feb. 2003.
[3] S. Borkar et al., Parameter variations and impact on circuits and microarchitecture, *IEEE Design Automation Conference*, pp. 338–342, 2003.

[4] Berkeley Predictive Technology Models, *http://www-device.eecs.berkeley.edu/∼ptm*.

[5] Y. Cao et al., New paradigm of predictive MOSFET and interconnect modeling for early circuit design, *Proceedings of the IEEE Custom Integrated Circuits Conference*, pp. 201–204, June 2000.

[6] Y. Cao et al., Design sensitivities to variability: extrapolations and assessments in nanometer VLSI, *IEEE International ASIC/SoC Conference*, pp. 411–415, Sept. 2002.

[7] S. R. Nassif, Design for variability in DSM technologies, *IEEE International Symposium on Quality Electronic Design*, pp. 451–454, 2000.

[8] C. Visweswariah, Death, taxes and failing chips, *IEEE Design Automation Conference*, pp. 343–347, 2003.

[9] K. A. Bowman, S. G. Duvall, and J. D. Meindl, Impact of die-to-die and within-die parameter fluctuations on the maximum clock frequency distribution, *IEEE International Solid-State Circuits Conference*, pp. 278–279, 2001.

[10] M. Eisele, J. Berthold, D. Schmitt-Landsiedel, and R. Mahnkopf, The impact of intra-die device parameter variations on path delays and on the design for yield of low voltage digital circuits, *IEEE Trans. VLSI Syst.*, Vol. 5, No. 4, pp. 360–368, Dec. 1997.

[11] D. Burnett, K. Erington, C. Subramanian, and K. Baker, Implications of fundamental threshold voltage variations for high-density SRAM and logic circuits, *IEEE Symposium on VLSI Technology*, pp. 15–16, 1994.

[12] Y. Cao et al., Yield optimization with energy-delay constraints in low-power digital circuits, *IEEE Conference on Electron Devices and Solid-State Circuits*, Hong Kong, Dec. 2003.

[13] S. Mukhopadhyay and K. Roy, Modeling and estimation of total leakage current in nano-scaled CMOS devices considering the effect of parameter variation, *IEEE International Symposium on Low Power Electronics and Design*, pp. 172–175, 2003.

[14] A. Srivastava, R. Bai, D. Blaauw, and D. Sylvester, Modeling and analysis of leakage power considering within-die process variations, *IEEE International Symposium on Low Power Electronics and Design*, pp. 64–67, 2002.

[15] H. Q. Dao, K. Nowka, and V. G. Oklobdzija, Analysis of clocked timing elements for dynamic voltage scaling effects over process parameter variation, *IEEE International Symposium on Low Power Electronics and Design*, pp. 56–59, 2001.

[16] S. Lin and C. K. Wong, Process-variation-tolerant clock skew minimization, *International Conference on Computer-Aided Design*, 1994.

[17] B. Gieseke et al., A 600 MHz superscalar RISC microprocessor with out-of-order execution, *IEEE International Solid-State Circuits Conference*, pp. 176–177, Feb. 1997.

[18] H. Ando et al., A 1.3 GHz fifth generation SPARC64 microprocessor, *IEEE International Solid-State Circuits Conference*, Feb. 2003.

[19] M. Bohr, Interconnect scaling: the real limiter to high performance ULSI, *Proceedings of the IEEE International Electron Devices Meeting*, pp. 241–244, Dec. 1995.

[20] K. Bernstein et al., *High Speed CMOS Design Styles*, Kluwer Academic, Norwell, MA, pp. 41–45, 1998.

[21] A. Kahng and M. Sarrafzadeh, Modern physical design: part V, tutorial, *International Conference on Computer-Aided Design*, Nov. 1999.

[22] D. Bailey and B. Benschneider, Clocking design and analysis for a 600-MHz alpha microprocessor, *IEEE J. Solid-State Circuits*, Vol. 33, No. 11, Nov. 1998.

[23] C. Bittlestone, A. Hill, V. Singhal, and N. V. Arvind, Architecting ASIC libraries and flows in nanometer era, *Design Automation Conference*, June 2003.

[24] K. Osada et al., Universal-V_{dd} 0.65–2.0 V 32 kB cache using voltage-adapted timing-generation scheme and a lithographical-symmetric cell, *IEEE International Solid-State Circuits Conference*, pp. 168–169, Feb. 2001.

[25] K. Bernstein, Design, process, and environmental contributors to CMOS delay variation, *SCCS near Limit Scaling Workshop*, 2003.

[26] A. Asenov et al., Increase in the random dopant induced threshold fluctuations and lowering in sub-100 nm MOSFETs due to quantum effects: a 3-D density-gradient simulation study, *IEEE Trans. Electron Devices*, Vol. 48, No. 4, Apr. 2001.

[27] P. Larsson, Measurements and analysis of PLL jitter caused by digital switching noise, *IEEE J. Solid-State Circuits*, Vol. 36, No. 7, July 2001.

[28] K. Osada et al., Universal-V_{dd} 0.65–2.0-V 32-kB cache using a voltage-adapted timing-generation scheme and a lithographically symmetrical cell, *IEEE J. Solid-State Circuits*, Vol. 36, No. 11, Nov. 2001.

[29] M. Yamaoka, K. Osada, and K. Ishibashi, 0.4-V logic library friendly SRAM array using rectangular-diffusion cell and delta-boosted-array-voltage scheme, *IEEE Symposium on VLSI Circuits*, 2002.

[30] D. Harris and M. A. Horowitz, Skew-tolerant domino circuits, *IEEE J. Solid-State Circuits*, Vol. 32, No. 11, Nov. 1997.

[31] G. A. Ruiz, Evaluation of three 32-bit CMOS adders in DCVS logic for self-timed circuits, *IEEE J. Solid-State Circuits*, Vol. 33, No. 4, Apr. 1998.

[32] L. G. Heller and W. R. Griffin, Cascode voltage switch logic: a differential CMOS logic family, *IEEE International Solid-State Circuits Conference*, pp. 16–17, 1984.

[33] K. Okada, Statistical modeling of device characteristics with systematic variability, *IEICE Trans. Fundam.*, Vol. E84-A, No. 2, Feb. 2001.

[34] M. J. M. Pelgrom, C. J. Duinmaijer, and A. P. G. Welbers, Matching properties of MOS transistors, *IEEE J. Solid State Circuits*, Vol. 24, No. 5, pp. 1433–1440, Oct. 1989.

[35] C. Michael and M. Ismail, Statistical modeling of device mismatch for analog MOS integrated circuits, *IEEE J. Solid State Circuits*, Vol. 27, No. 2, pp. 154–166, Feb. 1992.

[36] W. Zhang and Z. Yang, A new threshold voltage model for deep-submicron MOSFETs with nonuniform substrate dopings, *Microelectron. Reliab.*, Vol. 38, pp. 1465–1469, 1998.

本书作者

BAN P. WONG，担任了5年的IEEE国际固态电路会议技术程序委员会委员，并担任会议的主席、共同主席和小组会议的组织者。他拥有三项授权专利，并率领电路设计团队开发高性能、低功耗微处理器方法学与实施技术。他目前是NVIDIA公司的高级工程经理。ANURAG MITTAL，获得耶鲁大学应用物理博士。他参与了新型嵌入式NVM微处理器和微处理器解决方案的联合开发，包括世界上第一个与CMOS兼容的实用闪存技术。目前为Virage Logic公司的高级工程师。YU CAO，获得加州大学伯克利分校电气工程博士。现为伯克利无线研究中心的博士后研究人员，2000年被授予IEEE国际固态电路会议的Beatrice Winner奖。GREG STARR，获得亚利桑那州大学电气工程博士。目前是Xilinx公司的高级设计经理。

国际视野　科技前沿

国际信息工程先进技术译丛

《纳米CMOS电路和物理设计》
《现代通信原理》（原书第2版）
《认知无线电网络》
《高速数字系统的信号完整性和辐射发射》
《UMTS中的LTE：基于OFDMA和SC-FDMA的无线接入》
《生物医学工程学概论》（原书第2版）
《全面的功能验证：完整的工业流程》
《无线Mesh网络架构与协议》
《UMTS蜂窝系统的QoS与QoE管理》
《半导体制造与过程控制基础》
《WCDWA原理与开发设计》
《下一代移动系统：3G/B3G》
《IMS:IP多媒体概念和服务》（原书第2版）
《下一代无线系统与网络》
《深入浅出UMTS无线网络建模、规划与自动优化：理论与实践》
《HSDPA/HSUPA技术与系统设计——第三代移动通信系统宽带无线接入》
《无线传感器及元器件：网络、设计与应用》
《印制电路板——设计、制造、装配与测试》
《IPTV与网络视频：拓展广播电视的应用范围》
《多电压CMOS电路设计》
《微电子技术原理、设计与应用》
《蜂窝网络高级规划与优化2G/2.5G/3G/…向4G的演进》
《基于蜂窝系统的IMS——融合电信领域的VoIP演进》
《无线网络中的合作原理与应用》
《电生理学方法与仪器入门》
《移动电视：DVB-H、DMB、3G系统和富媒体应用》
《环境网络：支持下一代无线业务的多域协同网络》
《基于射频工程的UMTS空中接口设计与网络运行》
《未来UMTS的体系结构与业务平台：全IP的3G CDMA网络》
《UMTS-HSDPA系统的TCP性能》
《宽带无线通信中的空时编码》
《数字图像处理》（原书第4版）
《基于4G系统的移动服务技术》
《吉规模集成电路互连工艺及设计》
《高性能微处理器电路设计》

上架指导：工业技术/电子技术



ISBN 978-7-111-33083-7
封面设计：马精明

地址：北京市百万庄大街22号
电话服务
社服务中心：(010)88361066
销售一部：(010)68326294
销售二部：(010)88379649
读者购书热线：(010)88379203
邮政编码：100037
网络服务
门户网：http://www.cmpbook.com
教材网：http://www.cmpedu.com

封面无防伪标均为盗版



定价：98.00元